

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2004, 19:45:42 ; Search time 1836.19 Seconds  
(without alignments)  
11346.825 Million cell updates/sec

Title: US-10-066-007A-3  
Perfect score: 3969  
Sequence: 1 cggacgcagcctcgccagc.....ttcatgttgcgtacgtgcag 3969

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: Geneseqn1990s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3969	100.0	3969	3	AAA64473 DNA encod
2	353.2	8.9	1932	3	AAA64472 CDNA enco
C 3	60	1.5	2000	8	ADA71938 Rice gene
C 4	50.4	1.3	6741	3	AAAI0595 Gene enco
5	47	1.2	17993	6	ADA46721 Human tra
6	46.8	1.2	73465	6	ABQ88161 Human ost
7	43.6	1.1	919	10	ADC86904 Human GPC
C 8	42.2	1.1	32084	1	AAS29832 Human cyt
9	42.2	1.1	34680	4	AAK81232 Human inm
10	42.2	1.1	34680	4	AAK85743 Human inm
11	42	1.1	2000	8	ADA71938 Rice gene
C 12	41.4	1.0	21666	9	ADA02702 Human Nup
C 13	41.4	1.0	21666	10	ADB72440 Human CA
C 14	41.4	1.0	21666	10	ADP95950 Human hCG
C 15	41.2	1.0	12879	12	ADO48791 Human tis
C 16	41.2	1.0	13865	2	AAV40401 Human tis
C 17	41.2	1.0	13865	2	AAZ32165 Human cho
C 18	41.2	1.0	13865	8	ABZ71985 Human tis
C 19	40.8	1.0	4590	5	AAB24065 Yeast AOD
C 20	40.8	1.0	10732	3	AAAI0594 Gene enco
21	40.8	1.0	51365	9	ADA02990 Human CD2

22	40.8	1.0	51365	10	ADB72728
23	40.8	1.0	51365	10	ADC85470
24	40.8	1.0	51365	12	ADM74585
25	40.6	1.0	849	5	ASB89073
26	40.6	1.0	9642	6	ABL32356
27	39.8	1.0	18971	12	ADL08108
C 28	39.6	1.0	14041	4	AAH48024
29	39.2	1.0	1297	10	ADC87386
30	38.8	1.0	24801	10	ABZ20990
31	38.8	1.0	25301	12	ADO01541
32	38.8	1.0	41907	9	ADA37416
33	38.8	1.0	42998	8	ADA14747
34	38.8	1.0	42999	6	ASB65032
35	38.8	1.0	42999	10	AAE61411
36	38.6	1.0	473	3	AAAC43616
C 37	38.6	1.0	480	3	AAAC51536
C 38	38.6	1.0	957	3	AAAC34245
C 39	38.6	1.0	1368	4	AAAC60824
C 40	38.6	1.0	1425	5	ABV22247
C 41	38.6	1.0	1425	5	ABV28084
C 42	38.6	1.0	1425	5	ABV24599
C 43	38.6	1.0	170834	10	AAE62833
C 44	38.4	1.0	403	4	AAI81705
45	38.4	1.0	474	6	ABN62405

## ALIGNMENTS

RESULT 1  
AAA64473  
ID AAA64473 standard; DNA; 3969 BP.  
XX AC AAA64473;  
XX  
DT 15-SEP-2003 (revised)  
DT 02-JAN-2001 (first entry)  
XX  
DE DNA encoding an astaxanthin synthetase polypeptide.  
XX  
KW Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;  
KW antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.  
XX Xanthophyllomyces dendrorhous.  
XX  
FH Key Location/Qualifiers  
FT 5'UTR 517..518  
FT FT /\*tag= b  
FT CDS 535..3700  
FT FT /\*tag= a  
FT FT /\*product= "astaxanthin synthetase"  
FT FT 535..652  
FT FT /\*tag= c  
FT FT /\*number= 1  
FT intron 653..734  
FT FT /\*tag= d  
FT FT /\*number= 1  
FT exon 735..783  
FT FT /\*tag= e  
FT FT /\*number= 2  
FT intron 784..898  
FT FT /\*tag= f  
FT FT /\*number= 2  
FT exon 899..1015  
FT FT /\*tag= g  
FT FT /\*number= 3  
FT intron 1016..1087  
FT FT /\*tag= h  
FT FT /\*number= 3  
FT exon 1088..1179  
FT FT /\*tag= i  
FT FT /\*number= 4  
FT intron 1180..1302

```
FT FT /*tag= j
FT FT /number= 4
FT FT exon 1303. .1517
FT FT /*tag= k
FT FT /number= 5
FT FT intron 1518. .1600
FT FT /*tag= l
FT FT /number= 5
FT FT exon 1601. .1634
FT FT /*tag= m
FT FT /number= 6
FT FT intron 1635. .1723
FT FT /*tag= n
FT FT exon 1724. .1866
FT FT /*tag= o
FT FT /number= 7
FT FT intron 1867. .1939
FT FT /*tag= p
FT FT /number= 7
FT FT exon 1940. .1999
FT FT /*tag= q
FT FT /number= 8
FT FT intron 2000. .2081
FT FT /*tag= r
FT FT exon 2082. .2181
FT FT /*tag= s
FT FT /number= 9
FT FT intron 2182. .2257
FT FT /*tag= t
FT FT /number= 9
FT FT exon 2258. .2354
FT FT /*tag= u
FT FT /number= 10
FT FT intron 2355. .2431
FT FT /*tag= v
FT FT /number= 10
FT FT exon 2432. .2542
FT FT /*tag= w
FT FT /number= 11
FT FT intron 2543. .2618
FT FT /*tag= x
FT FT /number= 11
FT FT exon 2619. .2652
FT FT /*tag= y
FT FT /number= 12
FT FT intron 2653. .2742
FT FT /*tag= z
FT FT /number= 12
FT FT exon 2743. .2814
FT FT /*tag= aa
FT FT /number= 13
FT FT intron 2815. .2962
FT FT /*tag= ab
FT FT /number= 13
FT FT exon 2963. .3050
FT FT /*tag= ac
FT FT /number= 14
FT FT intron 3051. .3113
FT FT /*tag= ad
FT FT /number= 14
FT FT exon 3114. .3171
FT FT /*tag= ae
FT FT /number= 15
FT FT intron 3172. .3247
FT FT /*tag= af
FT FT /number= 15
FT FT exon 3248. .3321
FT FT /*tag= ag
FT FT /number= 16
FT FT intron 3322. .3398
FT FT /*tag= ah
FT FT /number= 16
FT FT

FT exon 3399. .3423
FT FT /*tag= ai
FT FT /number= 17
FT FT intron 3424. .3513
FT FT /*tag= aj
FT FT /number= 17
FT FT exon 3514. .3700
FT FT /*tag= ak
FT FT /number= 18
FT FT polyA_signal 3865. .3866
FT FT /*tag= al
FT FT
XX EP1035206-A1.
XX
XX 13-SEP-2000.
XX
XX 03-MAR-2000; 2000EP-00104430.
XX
XX 09-MAR-1999; 99EP-00104668.
XX 01-FEB-2000; 2000EP-00101666.
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX Hoshino T, Ojima K, Setoguchi Y;
XX
XX WPI; 2000-559874/52.
XX P-PSDB; AAB08713.
XX
XX Novel polynucleotide encoding astaxanthin synthase useful for producing
XX recombinant cells for producing astaxanthin from beta-carotene.
XX
XX Claim 4; Page 30-33; 46pp; English.
XX
XX The present sequence encodes an astaxanthin synthetase polypeptide of
XX Phaffia rhodozyma. The enzyme is involved in the last step of the
XX astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin. P.
XX rhodozyma is a carotenogenic yeast strain. The astaxanthin synthetase
XX polynucleotides and polypeptides are useful for producing astaxanthin.
XX Astaxanthin is an antioxidant which may be used to protect living cells
XX against diseases such as cancer. Astaxanthin is also used as a colouring
XX reagent, e.g. in farmed fish like salmon to impart an orange-red
XX coloration. (Updated on 15-SEP-2003 to standardise OS field)
XX
XX Sequence 3969 BP; 937 A; 942 C; 869 G; 1221 T; 0 U; 0 Other;

Query Match 100.0%; Score 3969; DB 3; Length 3969;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGACCGAAGCCTCGCCAGCAGTTGATCAAGCGAACCAGCCGAAACAATCTGCGCGGCC 60
DB 1 CGGACCGAAGCCTCGCCAGCAGTTGATCAAGCGAACCAGCCGAAACAATCTGCGCGGCC 60
QY 61 TGGAGGAGCGGAGCGGAGGAGCAGGTGATGCATCGGTGGAGAGATCAGTAGTG 120
DB 61 TGGAGGAGCGGAGCGGAGGAGCAGGTGATGCATCGGTGGAGAGATCAGTAGTG 120
QY 121 TGTGTATGTGTAGTGTAGTTGGGTGTGCCATGTGCTTCTTCTTATCATCATCATT 180
DB 121 TGTGTATGTGTAGTGTAGTTGGGTGTGCCATGTGCTTCTTCTTATCATCATCATT 180
QY 181 TCTTTAAATCTCTACATTTGAATGTTTACCGGAACGGGCTTTGATGATACGGACCAC 240
DB 181 TCTTTAAATCTCTACATTTGAATGTTTACCGGAACGGGCTTTGATGATACGGACCAC 240
QY 241 GTTGTGTAACCCAGTTCGATTGAGATTAGATAGCCGATCCGTCGATCAGATCTCG 300
DB 241 GTTGTGTAACCCAGTTCGATTGAGATTAGATAGCCGATCCGTCGATCAGATCTCG 300
QY 301 ATCTAGAGCGACATCTGGCTCGATCGGTCCTTTCGCCGAAATCAGGCGACCGATCAGGCA 360
DB 301 ATCTAGAGCGACATCTGGCTCGATCGGTCCTTTCGCCGAAATCAGGCGACCGATCAGGCA 360
```

Qy 361 GAGGAACGCGAGGCGGAACGAGACGACACACCATCATCATCAGCCATGCTCTTTTGT 420  
 Db 361 GAGGAACGCGAGGCGGAACGAGACGACACACCATCATCATCAGCCATGCTCTTTTGT 420  
 Qy 421 GATCGTTTTTACATACCTACCGGCGGATTCCTTTTCTTTCTCTCTGCGCATCTTT 480  
 Db 421 GATCGTTTTTACATACCTACCGGCGGATTCCTTTTCTTTCTCTCTGCGCATCTTT 480  
 Qy 481 GCATTCCTATCTCGTGAATGATCGATCCGATTCCTGCACTTCTTCCATATGTTT 540  
 Db 481 GCATTCCTATCTCGTGAATGATCGATCCGATTCCTGCACTTCTTCCATATGTTT 540  
 Qy 541 ATCTTGCTCTGCTCACAGGTGCTTTAGGCTCGCTGCTTTCTCATGGGCACTCCATAGG 600  
 Db 541 ATCTTGCTCTGCTCACAGGTGCTTTAGGCTCGCTGCTTTCTCATGGGCACTCCATAGG 600  
 Qy 601 TTCTTCAGTCTTTACCTCGCTCGAGGCGATCTTCACTGTATTAACCTTCAGGGTAAGAAT 660  
 Db 601 TTCTTCAGTCTTTACCTCGCTCGAGGCGATCTTCACTGTATTAACCTTCAGGGTAAGAAT 660  
 Qy 661 TGAGCTCTGGAATCATCTGTTGTAATCTTAATCTCACTCATCTTCTCTCTTTCT 720  
 Db 661 TGAGCTCTGGAATCATCTGTTGTAATCTTAATCTCACTCATCTTCTCTCTTTCT 720  
 Qy 721 CATCTCTCTTACGCGCGGAATCATACCACTACTTTACAGGCAATTTTTTACATCTT 780  
 Db 721 CATCTCTCTTACGCGCGGAATCATACCACTACTTTACAGGCAATTTTTTACATCTT 780  
 Qy 781 CTGCTGAGTTTTCATCATCTGCTCAGTCGTCCTTACAGTCACTCGCTACGACCTT 840  
 Db 781 CTGCTGAGTTTTCATCATCTGCTCAGTCGTCCTTACAGTCACTCGCTACGACCTT 840  
 Qy 841 TCGGACCGCTTCTTTCTATGTGAATCTGATCTTTGTTGTTTACGAGACACAGAG 900  
 Db 841 TCGGACCGCTTCTTTCTATGTGAATCTGATCTTTGTTGTTTACGAGACACAGAG 900  
 Qy 901 CTGCTACAGTGAAGAGCATCGGAGTACAGAGAAAATACGGAAGACCCCTCGGTTT 960  
 Db 901 CTGCTACAGTGAAGAGCATCGGAGTACAGAGAAAATACGGAAGACCCCTCGGTTT 960  
 Qy 961 CTGGGATCTGAGGACCGCTCTTGAATCTGAGCTCGACCGATCCGAAAGTCTTCAACAGTT 1020  
 Db 961 CTGGGATCTGAGGACCGCTCTTGAATCTGAGCTCGACCGATCCGAAAGTCTTCAACAGTT 1020  
 Qy 1021 TCCATCCGAACCCCTCATCTCTGCTGATCAATCACTGATGATTAACGCACTTTGAA 1080  
 Db 1021 TCCATCCGAACCCCTCATCTCTGCTGATCAATCACTGATGATTAACGCACTTTGAA 1080  
 Qy 1081 TGGACGTGTGATGAAGAGCCCTACGATCTCGGAAACCTGATGCGCGCTCGAGTGC 1140  
 Db 1081 TGGACGTGTGATGAAGAGCCCTACGATCTCGGAAACCTGATGCGCGCTCGAGTGC 1140  
 Qy 1141 TCAGAAATGCTACGGAGATGGTGTGTTTACGGCGGAAGTGTCTTTCAAGTTCTCTTAT 1200  
 Db 1141 TCAGAAATGCTACGGAGATGGTGTGTTTACGGCGGAAGTGTCTTTCAAGTTCTCTTAT 1200  
 Qy 1201 ATCACTCTAATCGACTCGGCGGATTTGAATCAACATTTCTACGAGCGCTGCACTTG 1260  
 Db 1201 ATCACTCTAATCGACTCGGCGGATTTGAATCAACATTTCTACGAGCGCTGCACTTG 1260  
 Qy 1261 TTTTCACTTCATGCTCTCGGTGATCTTGTCTCATCTCATAGGTGAAGCTCATAGCGAC 1320  
 Db 1261 TTTTCACTTCATGCTCTCGGTGATCTTGTCTCATCTCATAGGTGAAGCTCATAGCGAC 1320  
 Qy 1321 ATCGAAGGATCATGATCCCTCTCTGCTGCTGCTCAGGCGGTTAAGTGTGATGCTCCCAATTT 1380  
 Db 1321 ATCGAAGGATCATGATCCCTCTCTGCTGCTGCTCAGGCGGTTAAGTGTGATGCTCCCAATTT 1380  
 Qy 1381 TCTTAGAAAAGGATGTAAGCTTGTGCAAGAGATGATGGAAGATGCGGCTGCAAGGATA 1440  
 Db 1381 TCTTAGAAAAGGATGTAAGCTTGTGCAAGAGATGATGGAAGATGCGGCTGCAAGGATA 1440  
 Qy 1441 TGGCCGTGGGAGAGTCCGCCCGGTGAAAAAGGCAACACGACTCGAGACCGAAGGAGTGC 1500

1441 TGGCCGTGGGAGAGTCCGCCCGGTGAAAAAGGCAACCGACTCGAGACCGAAGGAGTGC 1500  
 1501 ATGTAAAAAGATTTGGTCTGTGAGTACCCGCTTATTCCTTCACTTGTATGAGCAAGCATAT 1560  
 1501 ATGTAAAAAGATTTGGTCTGTGAGTACCCGCTTATTCCTTCACTTGTATGAGCAAGCATAT 1560  
 1561 CAGGAAAGGTTCAATTTGACTGACAAACACTATCTTACAGGGTCCAGTCTCTGAGCGT 1620  
 1561 CAGGAAAGGTTCAATTTGACTGACAAACACTATCTTACAGGGTCCAGTCTCTGAGCGT 1620  
 1621 CATGCTCTTTCAGGTCAGTCTTACTCTCTTATAAATGCTCCACATATGATGATGTA 1680  
 1621 CATGCTCTTTCAGGTCAGTCTTACTCTCTCTTATAAATGCTCCACATATGATGATGTA 1680  
 1681 CTGACATCTCTTCTTATATTTGATACGACGCTCATATGTCCAGGATTTGATTAAGAGC 1740  
 1681 CTGACATCTCTTCTTATATTTGATACGACGCTCATATGTCCAGGATTTGATTAAGAGC 1740  
 1741 GACTCGCTCCAGAACCAAGCAACTGAGCTCTATGTTCGCTTTTGTCCGACTTTACCGATGGG 1800  
 1741 GACTCGCTCCAGAACCAAGCAACTGAGCTCTATGTTCGCTTTTGTCCGACTTTACCGATGGG 1800  
 1801 TTTGCTCTTACTTGGACTCGTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCCGA 1860  
 1801 TTTGCTCTTACTTGGACTCGTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCCGA 1860  
 1861 ACTATGTTATGTTCTCCCATTTCTTATATCAAAAGATTTATGATAGGTTTACTTGTAAAA 1920  
 1861 ACTATGTTATGTTCTCCCATTTCTTATATCAAAAGATTTATGATAGGTTTACTTGTAAAA 1920  
 1921 TTTTCACTTCTGTAACCAAGAACCGAGACATGAGATACCTTTTGACTCAAGGATTTAGCAGT 1980  
 1921 TTTTCACTTCTGTAACCAAGAACCGAGACATGAGATACCTTTTGACTCAAGGATTTAGCAGT 1980  
 1981 TTCCCGAGAGTTGGGATCGTAAGTCCAGATCAAGCTCTCTGATATTTCTTGGTCAATC 2040  
 1981 TTCCCGAGAGTTGGGATCGTAAGTCCAGATCAAGCTCTCTGATATTTCTTGGTCAATC 2040  
 2041 ATCTTAACTCTTAGGCTCATTTCACTCCATGCTGGGCAATAGGAGCTTATGGAGCAAAAGA 2100  
 2041 ATCTTAACTCTTAGGCTCATTTCACTCCATGCTGGGCAATAGGAGCTTATGGAGCAAAAGA 2100  
 2101 AGCAGGCGTGTGGCTCAGCTTCCGATCAGGCTGTTGATATAAAGAGATGTTCAAGTTC 2160  
 2101 AGCAGGCGTGTGGCTCAGCTTCCGATCAGGCTGTTGATATAAAGAGATGTTCAAGTTC 2160  
 2161 GGGATATCTTAAGTCTCTTAGTGTAGTAAAGCTTTTAAACGTATATACAGAGCGCGACA 2220  
 2161 GGGATATCTTAAGTCTCTTAGTGTAGTAAAGCTTTTAAACGTATATACAGAGCGCGACA 2220  
 2221 TTTTTCCTCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAAAATCCGCGCCAAC 2280  
 2221 TTTTTCCTCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAAAATCCGCGCCAAC 2280  
 2281 CTGCTGATCTCAAAAGCTGTCGATGAGGAGTCTCGCTCAGATCAGTAACTCTGTTA 2340  
 2281 CTGCTGATCTCAAAAGCTGTCGATGAGGAGTCTCGCTCAGATCAGTAACTCTGTTA 2340  
 2341 TTTTCTGGATGATGATGTTGTTCTTCTTCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 2400  
 2341 TTTTCTGGATGATGATGTTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2400  
 2401 ATAGAGTCTTTATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2460  
 2401 ATAGAGTCTTTATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2460  
 2461 TTTTCACTGCTTCAAGAACAAAGCCGTTCCAGGATTAACCTTCCGAGAGAAATTTGTGAGA 2520  
 2461 TTTTCACTGCTTCAAGAACAAAGCCGTTCCAGGATTAACCTTCCGAGAGAAATTTGTGAGA 2520  
 2521 TCGACACCGATATGCTTACGCTGAGGATGTTTGTGATGCTAAATTTACTTCTTCTTCTTCTTCT 2580

Db 2521 TCGACACGGATATGCCCTACGCTGTGAGGATGTTTTTGATGCTAAATTAATCTTCTCTGCA 2580  
 Qy 2581 AATGACTAAACGGCTTCCATCTTGATCCATTTTAGAGACGAATTAATGCGTTGCC 2640  
 Db 2581 AATGACTAAACGGCTTCCATCTTGATCCATTTTAGAGACGAATTAATGCGTTGCC 2640  
 Qy 2641 TAATCTCAAGCGGTGGTCTTCGATCTTGCTTCTGCTTCCAAATACAAATACGAGTAT 2700  
 Db 2641 TAATCTCAAGCGGTGGTCTTCGATCTTGCTTCTGCTTCCAAATACAAATACGAGTAT 2700  
 Qy 2701 TGCTCACTGATTTGCGCTACCGGCTGTGGAATTTAACTAGTTTCTTAAGGAGTCTCT 2760  
 Db 2701 TGCTCACTGATTTGCGCTACCGGCTGTGGAATTTAACTAGTTTCTTAAGGAGTCTCT 2760  
 Qy 2761 CGTCTAGACCGCTCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAAGACGTATG 2820  
 Db 2761 CGTCTAGACCGCTCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAAGACGTATG 2820  
 Qy 2821 TGGCTTCATCAGGCATAAATTTCAATTCATATTCCTTTGTACATACGACATACAGGCTGAC 2880  
 Db 2821 TGGCTTCATCAGGCATAAATTTCAATTCATATTCCTTTGTACATACGACATACAGGCTGAC 2880  
 Qy 2881 CGAGCTCAAAATTCGGGCTTCTCTCTGCTGCTTCTTTCTGCGCTTCTTATCTTCA 2940  
 Db 2881 CGAGCTCAAAATTCGGGCTTCTCTCTGCTGCTTCTTTCTGCGCTTCTTATCTTCA 2940  
 Qy 2941 CTTCAACCAAAATTTGTCAGATTCATCCACTGCGGCGCTGCTCATTTGCTGAGATG 3000  
 Db 2941 CTTCAACCAAAATTTGTCAGATTCATCCACTGCGGCGCTGCTCATTTGCTGAGATG 3000  
 Qy 3001 GTCGGTTCATCAACGAGTCCGATCAGAAAGGACGATGGTTCATCTTCGTAAGTTTC 3060  
 Db 3001 GTCGGTTCATCAACGAGTCCGATCAGAAAGGACGATGGTTCATCTTCGTAAGTTTC 3060  
 Qy 3061 CTTTATTTCTCTGCTCCGTAATAGTTTCTGATAGCGGACCAATTCAGCGTTGT 3120  
 Db 3061 CTTTATTTCTCTGCTCCGTAATAGTTTCTGATAGCGGACCAATTCAGCGTTGT 3120  
 Qy 3121 CAACATCAATCGTTCAAAAGTTCAATTTATGAGAGATGACAGAGAAATTCAGGTACAAATC 3180  
 Db 3121 CAACATCAATCGTTCAAAAGTTCAATTTATGAGAGATGACAGAGAAATTCAGGTACAAATC 3180  
 Qy 3181 GTTTTCTTTTAAAGCCAAATCGGTTGCTGATCGTAATGACCGGCTCTCTTTTAAATTC 3240  
 Db 3181 GTTTTCTTTTAAAGCCAAATCGGTTGCTGATCGTAATGACCGGCTCTCTTTTAAATTC 3240  
 Qy 3241 TCGAAGACCGGAGGTGGCTTGAGGACGTAACAGACTCGCTCAACAGTATTGAAGCAC 3300  
 Db 3241 TCGAAGACCGGAGGTGGCTTGAGGACGTAACAGACTCGCTCAACAGTATTGAAGCAC 3300  
 Qy 3301 CCTATGACACCGAGCGGTGATGTTTATGATTTTATGATTTTATGATTTTATGATTTTGA 3360  
 Db 3301 CCTATGACACCGAGCGGTGATGTTTATGATTTTATGATTTTATGATTTTATGATTTTGA 3360  
 Qy 3361 CGTTGACCTTCGGCTTCCCTGTTGTAATCCAGTATCTCTGACCCAGAGCTTG 3420  
 Db 3361 CGTTGACCTTCGGCTTCCCTGTTGTAATCCAGTATCTCTGACCCAGAGCTTG 3420  
 Qy 3421 CTTGTAAGTTTCTTCTCATCTGCGCTTAGCAGTATCCGATCAGCCATCTAGTTCTTTG 3480  
 Db 3421 CTTGTAAGTTTCTTCTCATCTGCGCTTAGCAGTATCCGATCAGCCATCTAGTTCTTTG 3480  
 Qy 3481 TACGATTTGTTCTGACTCTCGACTTTGCGAGTGGTGGAGTTTCTGCGCGAGAT 3540  
 Db 3481 TACGATTTGTTCTGACTCTCGACTTTGCGAGTGGTGGAGTTTCTGCGCGAGAT 3540  
 Qy 3541 GAAGGCTTCTTGTGTTGCTCACTCTCGCTCGGGTCCAGTTCCAGCCCATCATCTCATCC 3600  
 Db 3541 GAAGGCTTCTTGTGTTGCTCACTCTCGCTCGGGTCCAGTTCCAGCCCATCATCTCATCC 3600  
 Qy 3601 AGAGTACGAGCATACCTTGATCATTTCCGCTCGATCGATCGTTGGTAGAGAGGA 3660  
 Db 3601 AGAGTACGAGCATACCTTGATCATTTCCGCTCGATCGATCGTTGGTAGAGAGGA 3660

Qy 3661 GGGGTACAGATGCGTTTGCAGGTCAAGCGGTGCGAATGAGTTGATTCATATGTTAA 3720  
 Db 3661 GGGGTACAGATGCGTTTGCAGGTCAAGCGGTGCGAATGAGTTGATTCATATGTTAA 3720  
 Qy 3721 GAGAAGTTCTATATCTCGAATGCTGACTAGACCAATGCTCTTTTGTATCGATTGTT 3780  
 Db 3721 GAGAAGTTCTATATCTCGAATGCTGACTAGACCAATGCTCTTTTGTATCGATTGTT 3780  
 Qy 3781 TCTCATACCGGCGAGCGCTATGACTTTACGTCGTATCGTCGCTTGGACTCTCTT 3840  
 Db 3781 TCTCATACCGGCGAGCGCTATGACTTTACGTCGTATCGTCGCTTGGACTCTCTT 3840  
 Qy 3841 CTTACCTTATATATTTTCCATCCGTCGTATATTTGCTATACGACGCTCTGTGCTC 3900  
 Db 3841 CTTACCTTATATATTTTCCATCCGTCGTATATTTGCTATACGACGCTCTGTGCTC 3900  
 Qy 3901 AACTCAATATTACGCTCTTTCATGCTTCTGTGCTCCATAGATGATCTTCATGTTGT 3960  
 Db 3901 AACTCAATATTACGCTCTTTCATGCTTCTGTGCTCCATAGATGATCTTCATGTTGT 3960  
 Qy 3961 CGACTGCAG 3969  
 Db 3961 CGACTGCAG 3969

## RESULT 2

ABA64472  
 ID ABA64472 standard; cDNA; 1932 BP.

XX ABA64472;

XX 15-SEP-2003 (revised)

DT 02-JAN-2001 (first entry)

XX cDNA encoding an astaxanthin synthetase polypeptide.

KW Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;  
 antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.

XX Xanthophyllomyces dendrorhous.

FH Key Location/Qualifiers  
 CDS 33..1706

FT /\*tag= a

FT /\*product= "astaxanthin synthetase"

FT polyA\_signal 1871

XX /\*tag= b

PN EP1035206-A1.

XX 13-SEP-2000.

PF 03-MAR-2000; 2000EP-00104430.

PR 09-MAR-1999; 99EP-00104668.

PR 01-FEB-2000; 2000EP-00101866.

XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.

PA Hoshino T, Ojima K, Setoguchi Y;

XX WPI; 2000-559874/52.

DR P-FSDB; AAB08713.

XX Novel polynucleotide encoding astaxanthin synthase useful for producing

PT recombinant cells for producing astaxanthin from beta-carotene.

XX Claim 3; Page 24-27; 46pp; English.

XX The present sequence encodes an astaxanthin synthetase polypeptide of

CC Phaffia rhodozyma. The enzyme is involved in the last step of the

CC astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin. P.



CC rhodozyma is a carotenogenic yeast strain. The astaxanthin synthetase  
 CC polynucleotides and polypeptides are useful for producing astaxanthin.  
 CC Astaxanthin is an antioxidant which may be used to protect living cells  
 CC against diseases such as cancer. Astaxanthin is also used as a colouring  
 CC reagent, e.g. in farmed fish like salmon to impart an orange-red  
 CC coloration. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1932 BP; 495 A; 456 C; 481 G; 500 T; 0 U; 0 Other;

Query Match 8.9%; Score 353.2; DB 3; Length 1932;  
 Best Local Similarity 99.2%; Pred. No. 6.7e-93;  
 Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 3508 TCGCAGTGGTTGCGATTTCCTCGCGGAGATGAAGGCTTCTTGTGTCACCTCCG 3567  
 DB 1514 TTGCTTTGGTTGCGATTTCCTCGCGGAGATGAAGGCTTCTTGTGTCACCTCCG 1573  
 QY 3568 TCGGGTCCAGTTCGAGCCCATCTCTCATCCAGAGTACGACACATCACCCTTGATCAT 3627  
 DB 1574 TCGGGTCCAGTTCGAGCCCATCTCTCATCCAGAGTACGACACATCACCCTTGATCAT 1633  
 QY 3628 TTCCCGTCTCGAATCGTTGGTAGAGAGAGGGGTACCATCGTTTCAGGTCAA 3687  
 DB 1634 TTCCCGTCTCGAATCGTTGGTAGAGAGAGGGGTACCATCGTTTCAGGTCAA 1693  
 QY 3688 GCGGTCGAATGAGTTGATTCTTCATATGTTAAGAGAAGTCTATCTGAGAATGTGTG 3747  
 DB 1694 GCGGTCGAATGAGTTGATTCTTCATATGTTAAGAGAAGTCTATCTGAGAATGTGTG 1753  
 QY 3748 ACTAGACAAATGCTTCTTGTATCGAATTTGTTCTCATACCGGGCAGGCGCTATGACT 3807  
 DB 1754 ACTAGACAAATGCTTCTTGTATCGAATTTGTTCTCATACCGGGCAGGCGCTATGACT 1813  
 QY 3808 TCTACGTCGTCATCGCTCGCTGGAATCTCTTCTTACCTATATATTCCATCCG 3865  
 DB 1814 TCTACGTCGTCATCGCTCGCTGGAATCTCTTCTTACCTATATATTCCATCCG 1871

## RESULT 3

ADA71938/c  
 ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5263.

XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H; Chen W; Cooper B; Glazebrook J; Goff SA; Hou Y;

XX PI Katagiri F; Qian S; Tao Y; Whitham S; Xie Z; Zhu T; Zou G;

XX DR WPI; 2003-175290/17.

XX PT Identifying at least one gene involved in plant resistance or response to  
 XX PT pathogenic infection for conferring resistance or tolerance to a plant to  
 XX PT bacterial, fungal or viral infection by determining or detecting plant  
 XX PT gene expression.

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 1.5%; Score 60; DB 8; Length 2000;  
 Best Local Similarity 9.7%; Pred. No. 3.7e-06;  
 Matches 63; Conservative 301; Mismatches 286; Indels 2; Gaps 1;

QY 2372 TCTCTATCTTAGCTGATTAAAGCACTAATAGAGGCTTTATGTTTCTGTTGATCAG 2431  
 DB 891 TYKSWRRMYMTWKAWTMTTCMAKWNATGWATMMWRMYTYCYAMTCAKCYK 832  
 QY 2432 AACTTCTTCGACACTTTCACATGATGTTTCACCGACTCTCAGAAGACAAAGCGTTCA 2491  
 DB 831 MAMTKWTTWACAWRATSWRWAMAGRWKRYKMBAYWWRWRCWAGWAMKSYRW 772  
 QY 2492 GGATAAATCTCGAAGAAATTTGTTCAGATCGACACGATATGCTAGCTGCTGAGGATG 2551  
 DB 771 KWKYATRYKWMWMTMMWSRWKSYRMWSGMRMRWSAWRYCSRKCAKTYASSAR 712  
 QY 2552 TTTTGTGCTAAATTAATCTTCTTCTCAATGACTAAACGGCTTCCATCTTGTATCC 2611  
 DB 711 WTKAKSYRYRWRWYKRWYRYWRSCTMRMRKRWKRWKASWKNWYRGA 652  
 QY 2612 ATTTAGAGAACTTAATGCGTTGCTTATCTCGAAGCGGTGTTCTCGATCTTGG 2671  
 DB 651 RSMYKSYSCSAKCKKRYMTSSYMTGMYGMYSYKSMST--SKMSYMGKMTCTMYT 594  
 QY 2672 TCTTGTCTTCCAAATCAATACGATATTGCTCATCTGATTTCGGTCTACGGGCTGG 2731  
 DB 593 SMKGSTRSRMGWSGMSRMYMWWKWRKRYMYKWKCTWRRCMCYRWGYTWTTSR 534  
 QY 2732 AATTAACTAGTTTGAAGGAGTCTTCTGCTTAGACCTCTCTAGTCCGTATGCTAAAC 2791  
 DB 533 SRMYTGRYKARYTSKRYMYKYRYCWYVYGYMYKCSYMMRYGYCKACKCCYAMCW 474  
 QY 2792 GTGAATCTTAAGGATGAAGCGTATGTTGGCTTCATCAGCATATTTTCATTTCATA 2851  
 DB 473 KAAVSGMMWYRYKYSKWRMRMSTKYMWSMKYKCRSMKYGAKGYCGKMMYTCYGYMKW 414  
 QY 2852 TTCCTTTGTATACATACAGGCTGACGAGCTCAAAATTCGGCTTCTCTTCTGTCG 2911  
 DB 413 YTYGSKYKSRCKYMYRMYKGYMYMYYSAYSSMMTYVYKYKWKYKRGYMSWY 354  
 QY 2912 TCTTTTCTGGCTTCTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2971  
 DB 353 GSKYKXKXCTWYCYMKCMRCYRWKMRKXKXKXKXKXKXKXKXKXKXKXKXKXKX 294  
 QY 2972 CTTCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3023  
 DB 293 SMRTAGKWRMSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWSRWS 242

## RESULT 4

AAA10595/c

ID AAA10595 standard; DNA; 6741 BP.

XX AC AAA10595;

XX DT 29-JUN-2000 (first entry)

XX DE Gene encoding a subunit of cellulose synthase.



```

FT      /*tag= g
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(15915, G)
FT      /*tag= r
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
XX      WO200272831-A2.
XX      19-SEP-2002.
XX      15-JAN-2002; 2002WO-US000929.
XX      25-JAN-2001; 2001US-00768781.
XX      (PEKE ) PE CORP NY.
XX      Merkulov G, Guegler K, Brandon RC, Di Francesco V, Beasley EM;
XX      WPI; 2002-759846/82.
XX      P-PSDB; AAE29155, AAE29156.
XX      New isolated human transporter proteins and nucleic acids, useful for
XX      developing therapeutic or diagnostic compositions, particularly for
XX      developing human therapeutic agents that modulate transporter activity in
XX      cells or tissues.
XX      Claim 4; Page 79-85; 87pp; English.
XX      The present invention relates to novel human transporter proteins and
XX      polynucleotides encoding such proteins. Sequences of the invention are
XX      useful for developing human therapeutics and diagnostic compositions.
XX      These molecules are particularly useful as models for developing human
XX      therapeutic targets, identifying therapeutic proteins or serving as
XX      targets for the development of human therapeutic agents that modulate
XX      transporter activity in cells and tissues that express the transporter.
XX      The peptides are useful for raising antibodies or eliciting an immune
XX      response, as reagents in assays designed to quantitatively determine the
XX      levels of the protein (or its binding partner or ligand) in biological
XX      fluids or as markers for tissues in which the corresponding protein is
XX      preferentially expressed. They are also used in gene therapy. The agents
XX      which bind to transporter sequences are useful for treating transporter-
XX      related conditions that are specific for the subfamily of transporters
XX      that the peptides belong to, particularly in cells and tissues that
XX      express the transporter. The present sequence is a genomic DNA encoding
XX      human transporter proteins form 1 and form 2. Transporter gene is located
XX      on chromosome 23
XX      Sequence 17993 BP; 4883 A; 3793 C; 3765 G; 5527 T; 0 U; 25 Other;
XX      Query Match      1.2%; Score 47; DB 6; Length 17993;
XX      Best Local Similarity 51.2%; Pred. No. 0.096;
XX      Matches 110; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY      409  TGTCTTTTGTGATCGTTTACATACACTACCCGTCGATTAACCTTCTTCTTCTTCT 468
Db      12618 TTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 12677
QY      469  CTGCGCATCTTGCATCTCTATCTCGTGTACATGATCCGATCTTGCACCTACTTT 528
Db      12678 CTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTC 12737
QY      529  CTCATATGTCATCTGTGCTGCTCAGAGGTGCTTAGGCGCTGCTGCTTCTTCATGG 588
Db      12738 TTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 12797
QY      589  GCATCCATAGCGTTCCTCAGTCTTTACCTCGGCTCC 623
Db      12798 TTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 12832
XX      RESULT 6
XX      ABQ88161 standard; cDNA; 73465 BP.

```

```

XX      ABQ88161;
XX      18-SEP-2002 (first entry)
XX      Human osteoblast differentiation related cDNA SEQ ID NO 68.
XX      Human; osteoblast; stem cell differentiation; bone tissue deposition;
XX      osteoporosis; osteopathic; ss.
XX      Homo sapiens.
XX      WO200250301-A2.
XX      27-JUN-2002.
XX      18-DEC-2001; 2001WO-US048276.
XX      18-DEC-2000; 2000US-0255882P.
XX      24-APR-2001; 2001US-0285691P.
XX      (GENE-) GENE LOGIC INC.
XX      (PROC ) PROCTER & GAMBLE CO.
XX      Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
XX      Mertz L;
XX      WPI; 2002-557663/59.
XX      Use of genes and their expression profiles associated with osteoblast
XX      differentiation for screening modulators bone formation, for diagnosing
XX      or treating e.g. osteoporosis, or as markers for the differentiation
XX      process.
XX      Claim 1; SEQ ID NO 68; 78pp + Sequence Listing; English.
XX      The invention relates to genes and their expression profiles are used
XX      for: (a) screening modulators of precursor stem cell differentiation into
XX      osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
XX      deposition of bone tissue, abnormal rate of osteoblast formation or
XX      osteoporosis; or (c) treating or monitoring treatment of the conditions
XX      cited in (b), or monitoring the progression of bone tissue deposition.
XX      Specific conditions include postmenopausal osteoporosis, glucocorticoid
XX      osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
XX      induced abnormalities in bone formation or bone loss, conditions that
XX      involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
XX      skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
XX      or fibrous dysplasia. The present sequence is that of an osteoblast
XX      differentiation associated cDNA marker of the invention. Note: The
XX      sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences
XX      Sequence 73465 BP; 21307 A; 15130 C; 15840 G; 21188 T; 0 U; 0 Other;
XX      Query Match      1.2%; Score 46.8; DB 6; Length 73465;
XX      Best Local Similarity 51.4%; Pred. No. 0.25;
XX      Matches 108; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY      409  TGTCTTTTGTGATCGTTTACATACACTACCCGTCGATTAACCTTCTTCTTCTTCT 468
Db      63916 TTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 63975
QY      469  CTGCGCATCTTGCATCTCTATCTCGTGTAAATGATCGGATCTTGCACCTACTTT 528
Db      63976 CTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 64035
QY      529  CTCATATGTCATCTGTGCTGCTCAGAGGTGCTTAGGCGCTGCTGCTTCTTCATGG 588
Db      64036 CTCTTTTCTTCTTCTTCTGACAGGTTCTCACCTGTGTCCAGCTGAGTCAGTGGC 64095
QY      589  GCATCCATAGCGTTCCTCAGTCTTTTACCTC 618

```

D6 64096 ATGATTATAGCTATTCGACGCTTGAATC 64125

## RESULT 7

ADG6904  
ID ADC86904 standard; DNA; 919 BP.

AC ADC86904;

DT 01-JAN-2004 (first entry)

DE Human GPCR gene SEQ ID NO:1357.

XX ds; gene; human; GPCR;

KW guanosine triphosphate-binding protein coupled receptor; gene therapy.

XX Homo sapiens.

PN EP1270724-A2.

PD 02-JAN-2003.

XX 18-JUN-2002; 2002EP-00013517.

XX 18-JUN-2001; 2001JP-00246789.

XX (NAAB-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;

XX WPI; 2003-315783/31.

XX P-PSDB; ADC86905.

XX New polynucleotide, useful for preparing a composition for treating a

PT patient in need of increased or suppressed activity or expression of the

PT guanosine triphosphate-binding protein coupled receptor.

XX Claim 1; SEQ ID NO 1357; 28pp; English.

XX The invention relates to a novel polynucleotide encoding a guanosine

CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of

CC the invention may have a use in gene therapy. The polynucleotide and

CC polypeptide are useful for preparing a composition for treating a patient

CC in need of increased or suppressed activity or expression of the

CC guanosine triphosphate-binding protein coupled receptor. The

CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the

CC invention.

XX Sequence 919 BP; 167 A; 226 C; 154 G; 372 T; 0 U; 0 Other;

Query Match 1.1%; Score 43.6; DB 10; Length 919;

Best Local Similarity 52.8%; Pred. No. 0.16;

Matches 94; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 404 AGCATGCTTTTTCGATTCGTTTATACACTACCGTCGATCTACCTTCTTTCTTT 463

DB 264 AGCCCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 323

QY 464 CTTCCTTGCATCTTTCGATTCCTATCTCGTGAACATCGATCTTGCCACT 523

DB 324 TCTTCTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 383

QY 524 ACTTCTCCATGATTCATCTTGGCTTCTGCTACAGGTGCTTAGCCCTGGCTTT 581

DB 384 CTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTT 441

## RESULT 8

AAS29832/c

ID AAS29832 standard; DNA; 32064 BP.

XX

AC AAS29832;

XX 21-NOV-2001 (first entry)

Human cytoskeletal element-related polypeptide encoding genomic DNA #19.

Cytoskeletal element-related protein; human; mouse; rabbit; goat; horse;

KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;

KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;

KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;

KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplas;

KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;

KW cerebrovascular disorder; nervous system disorder; bacterial infection;

KW fungal infection; viral infection; ocular disorder; endocrine disorder;

KW gastrointestinal disorder; renal disorder; respiratory disorder;

KW wound healing; skin aging; organ transplantation; food preservative;

KW tissue regeneration; anti-infertility; food additive.

XX Homo sapiens.

XX WC200155168-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001331.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

XX 18-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226881P.

XX 22-AUG-2000; 2000US-0226888P.

XX 22-AUG-2000; 2000US-0227182P.

XX 23-AUG-2000; 2000US-0227009P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 06-SEP-2000; 2000US-0229513P.

XX 06-SEP-2000; 2000US-0230437P.

XX 06-SEP-2000; 2000US-0230438P.

XX 08-SEP-2000; 2000US-0231242P.

XX 08-SEP-2000; 2000US-0231243P.

XX 08-SEP-2000; 2000US-0231244P.

XX 08-SEP-2000; 2000US-0231413P.



AC AAK81232;  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36044.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytotatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001354.  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184654P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214866P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 08-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236378P.  
PR 29-SEP-2000; 2000US-0236388P.  
PR 29-SEP-2000; 2000US-0236389P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251020P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.







PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interference. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.

XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 1.1%; Score 42; DB 8; Length 2000;  
Best Local Similarity 10.3%; Pred. No. 0.78;  
Matches 89; Conservative 377; Mismatches 387; Indels 11; Gaps 3;

QY 1328 GATCATGATCCCTCTCTGCTCGCTCAGGCGGTAAAGTCGATGCTCCCAATTTCTTAGA 1387

DB 68 SASASGRTGSKSSGYSYSGKMKRYKRSKRWGRGRRGRMRMRMRVRRRCARSGR 127

QY 1388 AAAAGGTATGGAATTCGACAAAGATGATGAGGATGCGGCTCAGAAGATATGCGCCT 1447

DB 128 MAGSGRMMGKSRMSYMMWCYARGCGSKRKSXGSGNGKTCRERGAGSGWSGAKYK 187

QY 1448 GGGAGAGTCGGCGGTAAAAGAGGCAACGACGCTGAGACCGAAGAGGATGATGATAA 1507

DB 188 SGSNKRMMNMSCGSGCRSRSYRSYRGTYSRYGTYKMTYYSASRCMRAYNTTYSWA 247

QY 1508 GGATGGGTCTGAGTACCGCGCTATTCTCTTACCTTGATGGAGGATATCAAGGAA 1567

DB 248 CSSYTWCRSXRSRSMWKKRMRKRSRSGYSHYSXKMWCTAYKXYSRWCVWYGGW 307

QY 1568 AGGTTGATGATGACAAACACTATTCTACAGGCTGAGCTACTC-----TGGACGT 1620

DB 308 RGATRYMGRYMRMAMMYKMYWYRGYKMGKRWAGRMMSMRKSKACACVYVWRMW 367

QY 1621 CATGGCTCTGAGGTCAGTCTACTCTCTTATAAATGCTCCACATATGATGATGTA 1680

DB 368 RMTERRWAKSRTSRKRRKRWCRKRYKRWGRYSRMRSCRRARWRRKRSRGRWKG 427

QY 1681 CTGACATGCTCTTCTTATTCGATAG-ACGTCATATGTCAGGATTTGACTATAAGAG 1739

DB 428 CRGCTCRKSYGMRKWSKRWASKYKWMYSRWYRKKCSRTTWGKTGSGWMTWGR 487

QY 1740 CGACTCGCTCCAGAACAGACCAATGAGCTCTATGCTGCTTTGTCGGACTACCATGG 1799

DB 488 CRYKKSGRMKRKRKRWRGMRKRYMSARYTMYCARKYYSAAKARCYRKG 547

QY 1800 GTT---GCTCTACCTTGGACTCGTTCAGGCTATCATGCTGGGATTTGACTTACTT 1856

DB 548 GYVWAGWMMKRYKMYNYKMYWYKRYKSKSYSCMSYASCKSARKAGAKYCKSRM 607

QY 1857 CGAACTATGATGTCGCTCTTTGATATCCAAAGATATGATGATGATGCTTCTGCT 1916

DB 608 SAWSKSMRSRKRCKASKRSARAYMMGGMTSRSMRWSKSYCYWKRKWSKMSKTCW 667

QY 1917 AAAATTTCACTATCGTGAACAGAAAGGAGATGATGATGATGATGATGATGATGAT 1976

DB 668 MYMSKYTYAKYGYWYRANCMYRWYRYRYRYRYRYRYRYRYRYRYRYRYRYRY 727

QY 1977 CAGTTTCCGACGAGTGGGATCGTAAAGTCCGAGATCAAGCCCTCTCTGAATATCTTGT 2036

DB 728 RYNTSWYKCKSKWYSWYYSWMAKTKWRRRYATRRMMMYRYSMKYWYTWCTWNG 787

QY 2037 CATCATCTTAACCTCTAGCTCATTCATCGCTGCTGCGCAATAGAGGCTTATGGAGCA 2096

DB 788 YWYWRWYMKRYMYKCTKTYWTSATYWTGTWAAWMAKTKVRMGMTGAKTRGRARK 847  
QY 2097 AAGAAGCAGCGCTGCTTGGCTCAGCTCCGATCAGGCTGTGTATATAAAGAGATGTTCAA 2156  
DB 848 ARYMMWATWCATKRWMTKGAKWAMTWMAKAWKRYWWSWRAWYYYKTRIRYKTCW 907

QY 2157 GCTCGGATATCTCTAAGTCTCTTA 2180

DB 908 WKARWGSWAYWYMWKWSAKYWMW 931

RESULT 12

ADA02702/c

ID ADA02702 standard; DNA; 21666 BP.

XX ADA02702;

XX AC ADA02702;

XX DT 06-NOV-2003 (first entry)

XX DE Human Nupr1 carcinoma associated gene, SEQ ID NO:1220.

XX KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;

XX KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;

XX OS Homo sapiens.

XX PN WO2003057146-A2.

XX PD 17-JUL-2003.

XX PF 26-DEC-2002; 2002WO-US041414.

XX PR 26-DEC-2001; 2001US-00035832.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW;

XX DR WPI; 2003-587068/55.

XX PT New recombinant nucleic acid encoding carcinoma associated protein,

XX useful for preparing compositions for treating carcinomas.

XX PS Claim 1; SEQ ID NO 1220; 245pp; English.

XX CC The invention relates to recombinant carcinoma associated (CA) nucleic  
XX acid sequences from mouse and human (ADA01482-ADA03094), and to  
XX recombinant carcinoma associated proteins (CAP) encoded by them. The  
XX invention also encompasses expression vectors and host cells comprising a  
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically  
XX binds to the protein, and a biochip comprising CA nucleic acid or  
XX fragments thereof. The sequences of the invention were identified using  
XX oncogenic retroviruses, which insert into the genome of the host organism  
XX at random. Many of these do not carry transduced host oncogenes or  
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a  
XX direct consequence of the effects of proviral integration into host  
XX protooncogenes. The CA nucleic acid sequences can be used to diagnose  
XX carcinoma (especially breast cancer, prostate cancer, lymphoma or  
XX leukaemia) or a propensity to carcinoma by determination of the sequence  
XX of a CA gene, or by determination of CA gene expression in particular  
XX tissues. CA nucleic acids, proteins and antibodies are also useful as  
XX therapeutic agents and in screening and evaluating drug candidates. The  
XX present sequence represents a specifically claimed human CA nucleic acid  
XX sequence of the invention. Note: The complete sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 21666 BP; 5584 A; 5360 C; 5215 G; 5507 T; 0 U; 0 Other;

Query Match 1.0%; Score 41.4; DB 9; Length 21666;

Best Local Similarity 50.2%; Pred. No. 4.8;		
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;		
Qy	409	TGTCCTTTTTTGATCGCTTTTACATACACCCGTCGATTCCTAAACCTTCTTCTTCTTCT 468
Db	13480	TTCTCTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 13421
Qy	469	CTTGCCATCTTGCAATCTCTATCTCGTGTAACATGATCGGATCTTCGCCACTACTTT 528
Db	13420	CTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 13361
Qy	529	CTCCATATGTCATCTTGCTTGCTGCACAGTGTCTTAGGCGCTGGCTTCTTCATGG 588
Db	13360	CTCCTTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 13301
Qy	589	GCATCCATAGCGCTTCTTCAGTCT 611
Db	13300	TCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 13278
RESULT 13		
ADB72440/c		
ID	ADB72440 standard; DNA; 21666 BP.	
XX	AC	ADB72440;
XX	AC	
DT	04-DEC-2003 (first entry)	
XX	Human CA gene.	
XX		
KW	human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;	
KW	cancer; neoplasm; adenocarcinoma; sarcoma; gene.	
XX		
OS	Homo sapiens.	
XX		
PN	WO2003008593-A2.	
XX		
PD	30-JAN-2003.	
XX		
PF	26-DEC-2001; 2001WO-US051291.	
XX		
PR	02-MAR-2001; 2001US-00798586.	
PR	23-OCT-2001; 2001US-00004113.	
PR	08-NOV-2001; 2001US-00052482.	
PR	30-NOV-2001; 2001US-00997722.	
PR	20-DEC-2001; 2001US-00034650.	
XX		
PA	(SAGR-) SAGRES DISCOVERY.	
XX		
PI	Morris DW, Engelhard EK;	
XX		
DR	WPI; 2003-239337/23.	
XX		
PT	New recombinant nucleic acid, useful for treating carcinomas, lymphomas,	
PT	cancers, neoplasm, adenocarcinoma, or sarcomas.	
XX		
PS	Claim 1; SEQ ID NO 268; 2304pp; English.	
XX		
CC	The invention relates to a novel recombinant nucleic acid comprising a	
CC	nucleotide sequence selected from any of the 660 sequences fully defined	
CC	in the specification. A polynucleotide of the invention has cytostatic	
CC	activity, and may have a use in gene therapy, or in a vaccine. The	
CC	recombinant nucleic acids and polypeptides are useful for treating	
CC	carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and	
CC	sarcomas. The present sequence represents a human gene of the invention.	
XX		
SQ	Sequence 21666 BP; 5584 A; 5360 C; 5215 G; 5507 T; 0 U; 0 Other;	
Query Match 1.0%; Score 41.4; DB 10; Length 21666;		
Best Local Similarity 50.2%; Pred. No. 4.8;		
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;		
Qy	409	TGTCCTTTTTTGATCGCTTTTACATACACCCGTCGATTCCTAAACCTTCTTCTTCTTCT 468

Qy	2816	TATGTGGCTTCATACGAGCAATATTTTCATTTCATATTCCTTTGTGCATACGAGCATACAGG	2875
Db	8240	TTTATGATGATCCACCTCCACTTTATCAATAGTAAATATATTTTCTCTCCCTTATAATTC	8181
Qy	2876	CTGACGAGCTCAAAATTCGGCTTCCTCTCTGTGTCATTTTTCTGGCCCTTCTTATCT	2935
Db	8180	TTTACTCTCTCCCTCTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT	8121
Qy	2936	TCATTCTTCA	2945
Db	8120	TTTTTAGACA	8111

RESULT 15  
ADO48791/c

AC ADO48791;  
AC  
XX  
XX  
DT 12-AUG-2004 (first entry)  
XX  
XX Human tissue factor genomic DNA.  
DE  
DE  
XX  
XX Human; tissue factor; Gene; ds; antisense oligonucleotide;  
KW phosphorothioate linkage; 2'-O-methoxyethyl sugar moiety;  
KW 5-methylcytosine; hyperproliferative disorder; cancer; cytostatic.

OS  
XX  
XX  
HOMO sapiens.  
  
US2004102402-A1.  
XX  
XX  
XX  
PD PD  
27-MAY-2004.  
XX XX  
PF 22-NOV-2002; 2002US-00304098.  
XX XX  
FR 22-NOV-2002; 2002US-00304098.  
XX XX  
PA (ISIS-) ISIS PHARM INC.

PI  
XX  
DR WFI; 2004-399732/37.  
DR GENBANK; NT\_029227.  
XX  
PI New compound targeted to a nucleic acid molecule encoding tissue factor,  
PT useful in diagnosing and treating hyperproliferative disorder.

Claim 1; SEQ ID NO 4; 56pp; English.

The invention relates to a compound targeted to a nucleic acid molecule encoding the human tissue factor polypeptide. The compound is an antisense oligonucleotide that specifically hybridises with the nucleic acid and inhibits expression of the polypeptide. The antisense oligonucleotide comprises at least one modified internucleoside linkage, i.e. a phosphorothioate linkage, at least one modified sugar moiety, preferably a 2'-O-methoxyethyl sugar moiety, or at least one modified nucleobase comprising a 5-methylcytosine. The antisense compounds are useful for modulating the expression of the human tissue factor polypeptide and in preparation of a composition for treating hyperproliferative disorders, e.g. cancer. This sequence represents genomic DNA encoding the human tissue factor polypeptide of the invention.

SQ Sequence 12879 BP; 3466 A; 2687 C; 2979 G; 3747 T; 0 U; 0 Other;  
Query Match 1.0%; Score 41.2; DB 12; Length 12879;  
Best Local Similarity 51.1%; Pred.No. 4.1;  
Matches 97; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
2756 CTCCTGCTAGACCCCTCCTAGTCGGTATGTTAACCCTGAATGTAAAGCATCAACGC 2815  
DB CTCCTCCTCCTCCTCCTCCTCCTAGCTACTCAACATGAAGTAGCGAGATCAACACC 8241



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 01:23:48 ; Search time 326.21 Seconds  
(without alignments)  
8648.176 Million cell updates/sec

Title: US-10-066-007A-3  
Perfect score: 3969  
Sequence: 1 cggaccgaagcctcgacgcg.....ttcatgtttgtgcactgcag 3969

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3969	100.0	3969	3	US-09-518-386B-4
2	353.2	8.9	1932	1	US-09-518-386B-2
3	82.8	2.1	7218	1	US-08-332-463-14
4	46.8	1.2	832	4	US-09-621-976-2813
5	43	1.1	395	4	US-09-894-844-45
6	43	1.1	1141	4	US-09-806-708B-22
7	42	1.1	1141	4	US-09-806-708B-22
8	41.2	1.0	13865	3	US-09-009-217-11
9	41.2	1.0	13865	3	US-09-009-656-11
10	41.2	1.0	13865	4	US-09-054-272-11
11	40.4	1.0	289	3	US-09-007-005-17
12	40.4	1.0	289	3	US-09-244-796-17
13	39.4	1.0	7218	1	US-08-332-463-14
14	38.2	1.0	90541	4	US-09-759-359A-3
15	38.2	1.0	90541	4	US-10-207-373-3
16	37.8	1.0	474	4	US-09-621-976-18033
17	37.4	0.9	195	4	US-09-270-767-3604
18	37.4	0.9	195	4	US-09-270-767-18886
19	37.4	0.9	1563	4	US-09-527-073-1
20	37.2	0.9	1776	4	US-09-252-991A-9406
21	37.2	0.9	2583	4	US-09-252-991A-9340
22	36.8	0.9	521	3	US-07-998-289B-9
23	36.8	0.9	567	3	US-07-998-289B-2
24	36	0.9	1102	4	US-09-620-312B-928
25	35.6	0.9	376	2	US-08-623-906A-18
26	35.6	0.9	1923	4	US-09-134-000C-466
27	35.4	0.9	370	2	US-08-332-766A-8

Sequence 15639, A  
Sequence 996, App  
Sequence 15639, A  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 1655, Ap  
Sequence 1, Appli  
Sequence 38, Appli  
Sequence 7, Appli  
Sequence 6597, Ap  
Sequence 3036, A  
Sequence 14709, A  
Sequence 1230, A  
Sequence 14806, A  
Sequence 12861, A  
Sequence 3, Appli  
Sequence 269, App

US-09-518-386B-4  
; Sequence 4, Application US/09518386B  
; Patent No. 6365386  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/09/518,386B  
; CURRENT FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3969  
; TYPE: DNA  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (517)..(518)  
; NAME/KEY: Intron  
; LOCATION: (784)..(898)  
; NAME/KEY: Intron  
; LOCATION: (1016)..(1087)  
; NAME/KEY: Intron  
; LOCATION: (1180)..(1302)  
; NAME/KEY: Intron  
; LOCATION: (1518)..(1600)  
; NAME/KEY: Intron  
; LOCATION: (1635)..(1723)  
; NAME/KEY: Intron  
; LOCATION: (1867)..(1939)  
; NAME/KEY: Intron  
; LOCATION: (2000)..(2081)  
; NAME/KEY: Intron  
; LOCATION: (2182)..(2257)  
; NAME/KEY: Intron  
; LOCATION: (2355)..(2431)  
; NAME/KEY: Intron  
; LOCATION: (2543)..(2618)  
; NAME/KEY: Intron  
; LOCATION: (2653)..(2742)  
; NAME/KEY: Intron  
; LOCATION: (2815)..(2962)  
; NAME/KEY: Intron

#### ALIGNMENTS

RESULT 1  
US-09-518-386B-4  
; Sequence 4, Application US/09518386B  
; Patent No. 6365386  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/09/518,386B  
; CURRENT FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3969  
; TYPE: DNA  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (517)..(518)  
; NAME/KEY: Intron  
; LOCATION: (784)..(898)  
; NAME/KEY: Intron  
; LOCATION: (1016)..(1087)  
; NAME/KEY: Intron  
; LOCATION: (1180)..(1302)  
; NAME/KEY: Intron  
; LOCATION: (1518)..(1600)  
; NAME/KEY: Intron  
; LOCATION: (1635)..(1723)  
; NAME/KEY: Intron  
; LOCATION: (1867)..(1939)  
; NAME/KEY: Intron  
; LOCATION: (2000)..(2081)  
; NAME/KEY: Intron  
; LOCATION: (2182)..(2257)  
; NAME/KEY: Intron  
; LOCATION: (2355)..(2431)  
; NAME/KEY: Intron  
; LOCATION: (2543)..(2618)  
; NAME/KEY: Intron  
; LOCATION: (2653)..(2742)  
; NAME/KEY: Intron  
; LOCATION: (2815)..(2962)  
; NAME/KEY: Intron

```
; LOCATION: (3051)..(3113)
; NAME/KEY: intron
; LOCATION: (3172)..(3247)
; NAME/KEY: intron
; LOCATION: (3322)..(3398)
; NAME/KEY: intron
; LOCATION: (3424)..(3513)
; NAME/KEY: polyA site
; LOCATION: (3565)..(3866)
; NAME/KEY: intron
; LOCATION: (653)..(734)
US-09-518-386B-4

Query Match      100.0%; Score 3969; DB 3; Length 3969;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGACCGAAGCCTCGCCAGAGTTGATCAAGCGAAACAAAGCCGAA 60
Db 1 CGGACCGAAGCCTCGCCAGAGTTGATCAAGCGAAACAAAGCCGAA 60
Qy 61 TGGAGGAGCGGAGCGGAGGACGACAGGTGATCGGGTGGACAGATCAGTAGTG 120
Db 61 TGGAGGAGCGGAGCGGAGGAGGACGAGGTGATCGGGTGGACAGATCAGTAGTG 120
Qy 121 TGTGTGATGTGTGTAGTGTAGTGTGGTGTGCCATGTGCTTCTTATCATCATATT 180
Db 121 TGTGTGATGTGTGTAGTGTAGTGTGGTGTGCCATGTGCTTCTTATCATCATATT 180
Qy 181 TCCTTAAATCTCTACATTTGAATGTTTACCGGAACCGGGTTTGATGATACAGACCAAC 240
Db 181 TCCTTAAATCTCTACATTTGAATGTTTACCGGAACCGGGTTTGATGATACAGACCAAC 240
Qy 241 GTTGTGTAACAGTTTCGATTGAGATTAGATAGAGTACGATGCGATCGATCGATCG 300
Db 241 GTTGTGTAACAGTTTCGATTGAGATTAGATAGAGTACGATGCGATCGATCGATCG 300
Qy 301 ATCTAGAGGACATCTGGCTCGATTCGGTTCCTCGCGAAGAAATCAGGCGACCGATCAGGGCA 360
Db 301 ATCTAGAGGACATCTGGCTCGATTCGGTTCCTCGCGAAGAAATCAGGCGACCGATCAGGGCA 360
Qy 361 GAGGAACCGGAGCGCGAAACGAGACAGACACCATCATCATCAGCCATGCTTTTTTGT 420
Db 361 GAGGAACCGGAGCGCGAAACGAGACAGACACCATCATCATCAGCCATGCTTTTTTGT 420
Qy 421 GATCGTTTTTACATPACTACCCGTCGATTTAACTTTCTTTCTTTCTTTCTTTCTTTCTTT 480
Db 421 GATCGTTTTTACATPACTACCCGTCGATTTAACTTTCTTTCTTTCTTTCTTTCTTTCTTT 480
Qy 481 GCATTCCTACTCTGCTGTAACATCGATCGGATTCCTGCGACCTACTTTCTCCATATGTTTC 540
Db 481 GCATTCCTACTCTGCTGTAACATCGATCGGATTCCTGCGACCTACTTTCTCCATATGTTTC 540
Qy 541 ATCTTGTCTTGTCTACAGGTGCTTTAGGCGCTGGCTGCTTTCTCATGGGCATCCATPAGCG 600
Db 541 ATCTTGTCTTGTCTACAGGTGCTTTAGGCGCTGGCTGCTTTCTCATGGGCATCCATPAGCG 600
Qy 601 TTCTTACGCTTTTACCTCGCTCGAGCGGATCTTCACTGATTAACCTTCAGGCTAGAGAT 660
Db 601 TTCTTACGCTTTTACCTCGCTCGAGCGGATCTTCACTGATTAACCTTCAGGCTAGAGAT 660
Qy 661 TGAGCTCTGGAATCATGCTTGTGTAAATCCTATAATCTCATTCATCTTCTTTCTTTCTTT 720
Db 661 TGAGCTCTGGAATCATGCTTGTGTAAATCCTATAATCTCATTCATCTTCTTTCTTTCTTT 720
Qy 721 CATCTCTCTTACGCGCGGAATCATACCAACTACTTTTACAGGCAATTTTTTAGACATCCT 780
Db 721 CATCTCTCTTACGCGCGGAATCATACCAACTACTTTTACAGGCAATTTTTTAGACATCCT 780
Qy 781 CTCGTGAGTTTTCATCATTTGGCTCAGTCGTCCTTAATCTTAACGATCATCGCTAACGACCTT 840
Db 781 CTCGTGAGTTTTCATCATTTGGCTCAGTCGTCCTTAATCTTAACGATCATCGCTAACGACCTT 840
```

```
841 TCGGACCGCTTCTCTCTATGTGAATCTGATCTTTTGGTTTGTACGAGACACAGAG 900
841 TCGGACCGCTTCTCTCTATGTGAATCTGATCTTTTGGTTTGTACGAGACACAGAG 900
901 CTCCTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAATACGAAAGCAACCTCCGGTTTG 960
901 CTCCTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAATACGAAAGCAACCTCCGGTTTG 960
961 CTGGATCGCTGGAGCACCGCTCTTGAATCTCGACCGATCCGAAAGTCTTCAACCAAGTTTG 1020
961 CTGGATCGCTGGAGCACCGCTCTTGAATCTCGACCGATCCGAAAGTCTTCAACCAAGTTTG 1020
1021 TCCATCCGAACCCCTCATCTCTCTGCTGATCAATTCATCACTGATTAACGCACTTTGAA 1080
1021 TCCATCCGAACCCCTCATCTCTCTGCTGATCAATTCATCACTGATTAACGCACTTTGAA 1080
1081 TGGACAGTGTGATGAAAGAGCCCTACGACTATCCGAAACCTGGTATGCGCGCTCGAGTGC 1140
1081 TGGACAGTGTGATGAAAGAGCCCTACGACTATCCGAAACCTGGTATGCGCGCTCGAGTGC 1140
1141 TCAGAAATTGCTTACCGGAGATGGTGTGTTACGGCGGAAGGTGCTTTTCAAGTTCTCTTAT 1200
1141 TCAGAAATTGCTTACCGGAGATGGTGTGTTACGGCGGAAGGTGCTTTTCAAGTTCTCTTAT 1200
1201 ATCACTCTAATCCACTCGGCGGATTTGAATCAACTTCTGACGAGCGCTGCACCTTG 1260
1201 ATCACTCTAATCCACTCGGCGGATTTGAATCAACTTCTGACGAGCGCTGCACCTTG 1260
1261 TTTTCACTTCAATGCTCTCGGTGATCTTGTCTCATCTCATAGTGAAGCTCATAAGCGAC 1320
1261 TTTTCACTTCAATGCTCTCGGTGATCTTGTCTCATCTCATAGTGAAGCTCATAAGCGAC 1320
1321 ATCGAAGGATCATGATCCCTCTCTGTCGGTCAAGCGCTTAAGTCATGGTCCCAATTT 1380
1321 ATCGAAGGATCATGATCCCTCTCTGTCGGTCAAGCGCTTAAGTCATGGTCCCAATTT 1380
1381 TCTTAGAAAAAGGTATGGAATCTGTCACAAAGATGATGAGGATGCGGCTGAGAAAGGATA 1440
1381 TCTTAGAAAAAGGTATGGAATCTGTCACAAAGATGATGAGGATGCGGCTGAGAAAGGATA 1440
1441 TGGCGGTGGAGAGTCCGCGCGTGAAGAAAGCAACAGACTCGAGACCGAAGAGTGC 1500
1441 TGGCGGTGGAGAGTCCGCGCGTGAAGAAAGCAACAGACTCGAGACCGAAGAGTGC 1500
1501 ATGTAAAGGATTTGGTCTGAGTACCGCTTATCTTCTTCACTTGTAGGACGAGCATAT 1560
1501 ATGTAAAGGATTTGGTCTGAGTACCGCTTATCTTCTTCACTTGTAGGACGAGCATAT 1560
1561 CAAGAAAGGTTTCAATGACTGACAAACACTATCTTACAGGCTGAGCTACTCTGGACGT 1620
1561 CAAGAAAGGTTTCAATGACTGACAAACACTATCTTACAGGCTGAGCTACTCTGGACGT 1620
1621 CATGGCTCTTCAGGTGAGTCTCTCTTATAAATGCTCCACATATGATGTCATGTA 1680
1621 CATGGCTCTTCAGGTGAGTCTCTCTTATAAATGCTCCACATATGATGTCATGTA 1680
1681 CTGACATGCTCTTCTATATTCGATAGGAGTCAATATGTCAGGATTTGACTATAAGAGC 1740
1681 CTGACATGCTCTTCTATATTCGATAGGAGTCAATATGTCAGGATTTGACTATAAGAGC 1740
1741 GACTCGCTCCAGAACAGACCAATGAGCTCTATGCTCTTTTTCGAGCTTACCGATGGG 1800
1741 GACTCGCTCCAGAACAGACCAATGAGCTCTATGCTCTTTTTCGAGCTTACCGATGGG 1800
1801 TTTGCTCTTACCTTGGACTCGTCAAGCTATCATGTGGGATTTGTACTTCTTCCGA 1860
1801 TTTGCTCTTACCTTGGACTCGTCAAGCTATCATGTGGGATTTGTACTTCTTCCGA 1860
1861 ACTATGTAATCTGCTGCTTCTTTTATATCAAAAGATATGATAGTGTACTTCTGCTAAAA 1920
1861 ACTATGTAATCTGCTGCTTCTTTTATATCAAAAGATATGATAGTGTACTTCTGCTAAAA 1920
1921 TTTTCACTTATCGTGAACAGAAACGAGACATGAGATACCTTTTGTACTCAAGATTTAGCAGT 1980
```



Db 1921 TTTTCACTATCGTGAACAGAAACGGAGACATGAGATACCTTTGACTCAAGGATTAGCAGT 1980  
Qy 1981 TTCCCGACGAGTTGGGATCGTAAGTGGCAGATCAAGCCCTCTCTGAATATTTCTTTGGTCAATC 2040  
Db 1981 TTCCCGACGAGTTGGGATCGTAAGTGGCAGATCAAGCCCTCTCTGAATATTTCTTTGGTCAATC 2040  
Qy 2041 ATCTTAACCTCTTAGGCTCATTTCAATCATGTGTGGCAATAGGAGCTTATCGAGCAAAAGA 2100  
Db 2041 ATCTTAACCTCTTAGGCTCATTTCAATCATGTGTGGCAATAGGAGCTTATCGAGCAAAAGA 2100  
Qy 2101 AGCAGGCGGTGTTGGCTTCAGCTTCCGATCAGGCTGTGTGATAAAAGGATGTTCAAGGTC 2160  
Db 2101 AGCAGGCGGTGTTGGCTTCAGCTTCCGATCAGGCTGTGTGATAAAAGGATGTTCAAGGTC 2160  
Qy 2161 GGGATATCCTAAGTCTCCTAGGTTAGTAACGTTTTTAAACGTTATATACAGAGCGGCGACA 2220  
Db 2161 GGGATATCCTAAGTCTCCTAGGTTAGTAACGTTTTTAAACGTTATATACAGAGCGGCGACA 2220  
Qy 2221 TTTCTTCCCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAACATCGCGGCCAAC 2280  
Db 2221 TTTCTTCCCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAACATCGCGGCCAAC 2280  
Qy 2281 CTGCTGAATCTCAAAAAGCTGTCCGATGAGGAGGTACTCGCTCAGATCAAGTAACTGTTA 2340  
Db 2281 CTGCTGAATCTCAAAAAGCTGTCCGATGAGGAGGTACTCGCTCAGATCAAGTAACTGTTA 2340  
Qy 2341 TTTGCTGGATATGAGTGTGATCCTTTCCCTCTCTATCCTTAGCTGATTAAAGGACACTA 2400  
Db 2341 TTTGCTGGATATGAGTGTGATCCTTTCCCTCTCTATCCTTAGCTGATTAAAGGACACTA 2400  
Qy 2401 ATAGAGTCTTTATGTTTCTGTTTGTATCAGAACTTCTTCGACAGCTTTGACATGGATGT 2460  
Db 2401 ATAGAGTCTTTATGTTTCTGTTTGTATCAGAACTTCTTCGACAGCTTTGACATGGATGT 2460  
Qy 2461 TTCAACGACACTCAGAGAGCAAAAGCGCTTCAGATCAAACTTCGAGAGAAATTTGTCAGA 2520  
Db 2461 TTCAACGACACTCAGAGAGCAAAAGCGCTTCAGATCAAACTTCGAGAGAAATTTGTCAGA 2520  
Qy 2521 TCGACACGGATATGCTTACGCTGTGAGGATGTTTTTGTATGCTAAATTAATCTTCTTGCA 2580  
Db 2521 TCGACACGGATATGCTTACGCTGTGAGGATGTTTTTGTATGCTAAATTAATCTTCTTGCA 2580  
Qy 2581 AATGACTTAAACGGCTTCCATTTCTGATCCATTTTAGACACAACTTAATGCGTTGCCCT 2640  
Db 2581 AATGACTTAAACGGCTTCCATTTCTGATCCATTTTAGACACAACTTAATGCGTTGCCCT 2640  
Qy 2641 TATCTCGAAGCGGTTGGTTCTCGATTTCTGGTCTTGTCTTCCAAATACAAATACGGATTAT 2700  
Db 2641 TATCTCGAAGCGGTTGGTTCTCGATTTCTGGTCTTGTCTTCCAAATACAAATACGGATTAT 2700  
Qy 2701 TGCTCATCTGATTTGGCTCTACGGGCTGTGGAAATTTAACTAGTTTGTAAAGGAGTCTCTT 2760  
Db 2701 TGCTCATCTGATTTGGCTCTACGGGCTGTGGAAATTTAACTAGTTTGTAAAGGAGTCTCTT 2760  
Qy 2761 CGTCTAGACCTCTCTAGTCCGATGCTTAACCGGTAATGCTTAAAGATGAAGACGATGT 2820  
Db 2761 CGTCTAGACCTCTCTAGTCCGATGCTTAACCGGTAATGCTTAAAGATGAAGACGATGT 2820  
Qy 2821 TGGCTTCATCAGCATAAATTTTCAATTTCAATTTCTTTGTACATACGATACAGGCTGAC 2880  
Db 2821 TGGCTTCATCAGCATAAATTTTCAATTTCAATTTCTTTGTACATACGATACAGGCTGAC 2880  
Qy 2881 CGAGCTCAAAATTCGGGCTTCTCTGCTGCTTCTTTTCTGGCCCTTCTATCTTCATT 2940  
Db 2881 CGAGCTCAAAATTCGGGCTTCTCTGCTGCTTCTTTTCTGGCCCTTCTATCTTCATT 2940  
Qy 2941 CTTCAACCAAAAATTTGTACAGTTTCATCCCACTTGGCGAGCCCTGTCAATGGTCGAGATGG 3000  
Db 2941 CTTCAACCAAAAATTTGTACAGTTTCATCCCACTTGGCGAGCCCTGTCAATGGTCGAGATGG 3000  
Qy 3001 GTGGTTCATCAACGAGTCCGGATTCAGAAAGGAAACGATGCTCATCTTCGTAAGTTTTC 3060

Db 3001 GTGGTTCATCAACGAGTCCGGATCACGAAGGAACGATGGTCATGCTTCGTAAGTTTTC 3060  
Qy 3061 CTTTATTTTCAATCTCTCCATGAAATAGTTTCTGTAGACGGGACCAATTCAGCGTTGTT 3120  
Db 3061 CTTTATTTTCAATCTCTCCATGAAATAGTTTCTGTAGACGGGACCAATTCAGCGTTGTT 3120  
Qy 3121 CAACATCAATCGTTCAAAAGTTTCATTTATGGAAGATGCAGAAGAAATTCAGGTACAATTC 3180  
Db 3121 CAACATCAATCGTTCAAAAGTTTCATTTATGGAAGATGCAGAAGAAATTCAGGTACAATTC 3180  
Qy 3181 GTTTTCTTTTAAAGCCCAATCGGTTTCTGTATCGTAATTTGACCGGCTCTCTTTTAAATTC 3240  
Db 3181 GTTTTCTTTTAAAGCCCAATCGGTTTCTGTATCGTAATTTGACCGGCTCTCTTTTAAATTC 3240  
Qy 3241 TCGAAGACCCGAGAGGTGGCTTGAGGAGCTAAACAGACTCGCTCAACAGTATTAAGGAC 3300  
Db 3241 TCGAAGACCCGAGAGGTGGCTTGAGGAGCTAAACAGACTCGCTCAACAGTATTAAGGAC 3300  
Qy 3301 CCTATGGACACCAGGCGAGCTGTATGTTTATGTAATTTTATCTTTGTGCAATTTTGC AAAA 3360  
Db 3301 CCTATGGACACCAGGCGAGCTGTATGTTTATGTAATTTTATCTTTGTGCAATTTTGC AAAA 3360  
Qy 3361 CGTTGAACTTTCGGCTTCCCTTGTGTTGAAATCCCACTTATCTCTGACCCAGAGCTTG 3420  
Db 3361 CGTTGAACTTTCGGCTTCCCTTGTGTTGAAATCCCACTTATCTCTGACCCAGAGCTTG 3420  
Qy 3421 CTTGTAAGTTTCTCTCATCTGGGCTTTAGCAGTATCCGATCAGCCATCTAGTCTTTTG 3480  
Db 3421 CTTGTAAGTTTCTCTCATCTGGGCTTTAGCAGTATCCGATCAGCCATCTAGTCTTTTG 3480  
Qy 3481 TACGATGTTTCTGACTCTCTCGACTTTCGAGTGTGGGATTTGCTGCGCGAGAT 3540  
Db 3481 TACGATGTTTCTGACTCTCTCGACTTTCGAGTGTGGGATTTGCTGCGCGAGAT 3540  
Qy 3541 GAAGGCTTCTTGTGTTGTCACCTCCGTCGGGTCAGTTTCGAGCCCATCTCTCATCC 3600  
Db 3541 GAAGGCTTCTTGTGTTGTCACCTCCGTCGGGTCAGTTTCGAGCCCATCTCTCATCC 3600  
Qy 3601 AGAGTACGAGCACATCACCTTTGATCANTTCCGCTCTCGAATCGTTGGTAGAGAGAAGGA 3660  
Db 3601 AGAGTACGAGCACATCACCTTTGATCANTTCCGCTCTCGAATCGTTGGTAGAGAGAAGGA 3660  
Qy 3661 GGGGTACAGATCGTTTTCGAGGTCAAGCCCGTTCGAATGAGTTGATTTTCATATGTAA 3720  
Db 3661 GGGGTACAGATCGGTTTTCGAGGTCAAGCCCGTTCGAATGAGTTGATTTTCATATGTAA 3720  
Qy 3721 GAGAAGTTCTATATCTGAGAATGTGACTAGGACAATGCCTTCTTTGTATCGATTTGTT 3780  
Db 3721 GAGAAGTTCTATATCTGAGAATGTGACTAGGACAATGCCTTCTTTGTATCGATTTGTT 3780  
Qy 3781 TCTCATACCCGGCAGCGGCTATGACTTCTACGTCGCTATFCGTCGCTCTGGACTCTCTT 3840  
Db 3781 TCTCATACCCGGCAGCGGCTATGACTTCTACGTCGCTATFCGTCGCTCTGGACTCTCTT 3840  
Qy 3841 CTTACCCCTATATATTTTCCATCCGCTCTGTATATTTGTCATACAGAGCTGTGTGCTC 3900  
Db 3841 CTTACCCCTATATATTTTCCATCCGCTCTGTATATTTGTCATACAGAGCTGTGTGCTC 3900  
Qy 3901 AACTCAATATTCAGCCTCTTTCATGCTCTGTGCTCTCCATAGATGTGATCTTCAATTTGT 3960  
Db 3901 AACTCAATATTCAGCCTCTTTCATGCTCTGTGCTCTCCATAGATGTGATCTTCAATTTGT 3960  
Qy 3961 CGACTGCGAG 3969  
Db 3961 CGACTGCGAG 3969

## RESULT 2

US-09-518-386B-2  
; Sequence 2, Application US/09518386B  
; Patent No. 6355386  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo

```

; APPLICANT: OJIMA, Kazuyuki
; APELLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE
; CURRENT APPLICATION NUMBER: US/09/518,386B
; CURRENT FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: EP 99104668.1
; PRIOR FILING DATE: 1999-03-09
; PRIOR APPLICATION NUMBER: EP 00101666.6
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (33)..(1706)
; NAME/KEY: polyA_site
; LOCATION: (1871)
; NAME/KEY: mRNA
; LOCATION: (14)..(1891)
; US-09-518-386B-2

Query Match      8.9%; Score 353.2; DB 3; Length 1932;
Best Local Similarity 99.2%; Pred. No. 2.8e-103;
Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3508 TGCAGTGGTTGGCGATTTCCTGTGCCGAGATGAAGGCTTCTTTGTTTGTCATCTCCG 3567
Db 1514 TTGCTTTGGTTGGCGATTTCCTGTGCCGAGATGAAGGCTTCTTTGTTTGTCATCTCCG 1573
QY 3568 TGGGTCAGTTCGAGGCCATCATCTCATCCAGAGTACGAGCACATCACCTTGATCAT 3627
Db 1574 TGGGTCAGTTCGAGGCCATCATCTCATCCAGAGTACGAGCACATCACCTTGATCAT 1633
QY 3628 TTCCCGTCTCTCGAATCGTTGGTAGAGAGAGAGGGGTACCCAGATCCGTTTGCAGGTCAA 3687
Db 1634 TTCCCGTCTCTCGAATCGTTGGTAGAGAGAGAGGGGTACCCAGATCCGTTTGCAGGTCAA 1693
QY 3689 GCGGTGCAATGAGTTGATTTCTATGTTAGAGAGGTTCTATATCTGAGATGTGTG 3747
Db 1694 GCGGTGCAATGAGTTGATTTCTATGTTAGAGAGGTTCTATATCTGAGATGTGTG 1753
QY 3748 ACTAGACAAATGCGCTTCTTTGATCGAGTTGTTTCTCATACCCGGCGAGGGGCTATGACT 3807
Db 1754 ACTAGACAAATGCGCTTCTTTGATCGAGTTGTTTCTCATACCCGGCGAGGGGCTATGACT 1813
QY 3808 TCTAGTCGTCTATCGTGGCTGTGAGATCTCTTCTTACCCCTATATATTATTCATCCG 3865
Db 1814 TCTAGTCGTCTATCGTGGCTGTGAGATCTCTTCTTACCCCTATATATTATTCATCCG 1871

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```









```

US-08-232-463-14

Query Match      1.0%; Score 39.4; DB 1; Length 7218;
Best Local Similarity 10.6%; Pred. No. 0.52;
Matches 19; Conservative 97; Mismatches 63; Indels 0; Gaps 0;

Qy      1344 CTGTCCCGCTCAGGCCGTTAAGTCGATGGTCCCAATTTCTTAGAAAAAGATATGGAACCTT 1403
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1475 CTATCTATGCAAGTAGTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRRRRRRRRR 1416

Qy      1404 GTCCACAAGATGATGAGGATCGCGGTGAGAAGCATATGCCCGTGGGAGAGTCGCGCGGT 1463
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy      1464 GAAAAGAGCAACCAAGCTCGAGACCGAAGGAGTCGATGTAAAGGATTTGGGTCGTGAG 1522
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297

RESULT 14
US-09-759-359A-3
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:

```

```

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001043
; CURRENT APPLICATION NUMBER: US/09/759,359A
; CURRENT FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 90541
; TYPE: DNA
; ORGANISM: Human
US-09-759-359A-3

Query Match          1.0%; Score 38.2; DB 4; Length 90541;
Best Local Similarity 47.7%; Pred. No. 9.9;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 414 TTTTGTGATCGTTTTACATACCTACCGTCGATTCCTAACTTCTTTCTTCTCTCTCGC 473
Db 34847 TTTCTTTCTTCTCTTTCTTTTCTTTCTTCTCTCTTCTTCTTCTTCTTCTTCTT 34900

QY 474 CACTTTTGCATCTCTATCTCGTGAACATCGATCGGATCTTGGCCACCTACTTTCTCCA 533
Db 34907 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 34960

QY 534 TAGTTTCATCTGGTCTTCTGTACAGGTGCTTTAGCCCTGGTCTTTCTCATGGGCATC 593
Db 34967 TCCTTTTCTCTCCCTCTCTCTCTCCCTCCCTCCCTCCCTCCCTCTTCTTCTTCTT 35020

QY 594 CATAGGGTCTTTTCAGTCTTTACCTCTCGTCGAGGGCATCTTCACTGTATAAAGCTT 648
Db 35027 CCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTCTTCTTCT 35081

RESULT 15
US-10-207-973-3
; Sequence 3, Application US/10207973
; Patent No. 6753175
; GENERAL INFORMATION:
; APPLICANT: ABU-THREIDEH, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001043
; CURRENT APPLICATION NUMBER: US/10/207,973
; CURRENT FILING DATE: 2002-07-31
; NUMBER OF SEQ ID NOS: 3

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 01:52:58 ; Search time 1935.06 Seconds  
(without alignments)  
11308.754 Million cell updates/sec

Title: US-10-066-007A-3  
Perfect score: 3969  
Sequence: 1 cggacgagcctgcgcagc.....ttcatgtttgtcgactgag 3969

Scoring table: IDENTITY NUC  
Gapop 10.0, Capext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	ID Description
1	3969	100.0	3969 14 US-10-066-007-4 Sequence 4, Appli
2	353.2	8.9	1932 14 US-10-066-007-2 Sequence 2, Appli
3	48	1.2	1849 13 US-10-027-632-264440 Sequence 264440,
4	48	1.2	1849 13 US-10-027-632-264441 Sequence 264441,
5	48	1.2	1849 15 US-10-027-632-264440 Sequence 264440,
6	48	1.2	1849 15 US-10-027-632-264441 Sequence 264441,
7	47	1.2	17933 9 US-09-768-781-5 Sequence 5, Appli
8	46.8	1.2	1849 13 US-10-027-632-264438 Sequence 264438,
9	46.8	1.2	1849 13 US-10-027-632-264439 Sequence 264439,
10	46.8	1.2	1849 15 US-10-027-632-264438 Sequence 264438,
11	46.8	1.2	1849 15 US-10-027-632-264439 Sequence 264439,
12	46.8	1.2	73465 17 US-10-450-826-68 Sequence 68, Appli

# ALIGNMENTS

RESULT 1  
US-10-066-007-4  
; Sequence 4, Application US/10066007  
; Publication No. US20030077691A1  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/10/066.007  
; CURRENT FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US/09/518,386  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3969  
; TYPE: DNA  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (517)..(518)  
; NAME/KEY: intron  
; LOCATION: (784)..(898)  
; NAME/KEY: intron  
; LOCATION: (1016)..(1087)  
; NAME/KEY: intron  
; LOCATION: (1180)..(1302)  
; NAME/KEY: intron  
; LOCATION: (1518)..(1600)

Sequence 168155,  
Sequence 168156,  
Sequence 168157,  
Sequence 168158,  
Sequence 168155,  
Sequence 168156,  
Sequence 168157,  
Sequence 168158,  
Sequence 1357, Ap  
Sequence 1701, Ap  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 162899,  
Sequence 162899,  
Sequence 117, App  
Sequence 137448,  
Sequence 137448,  
Sequence 208, App  
Sequence 459, App  
Sequence 23425, A  
Sequence 4, Appli  
Sequence 11, Appl  
Sequence 331, App  
Sequence 42, Appl  
Sequence 22067, A  
Sequence 356, App  
Sequence 95, Appl  
Sequence 329, App  
Sequence 18301, A  
Sequence 18301, A  
Sequence 130360,  
Sequence 130362,

```
NAME/KEY: intron
LOCATION: (1835)..(1723)
NAME/KEY: intron
LOCATION: (1867)..(1939)
NAME/KEY: intron
LOCATION: (2000)..(2081)
NAME/KEY: intron
LOCATION: (2182)..(2257)
NAME/KEY: intron
LOCATION: (2355)..(2431)
NAME/KEY: intron
LOCATION: (2543)..(2618)
NAME/KEY: intron
LOCATION: (2653)..(2742)
NAME/KEY: intron
LOCATION: (2815)..(2962)
NAME/KEY: intron
LOCATION: (3051)..(3113)
NAME/KEY: intron
LOCATION: (3172)..(3247)
NAME/KEY: intron
LOCATION: (3322)..(3398)
NAME/KEY: intron
LOCATION: (3424)..(3513)
NAME/KEY: polyA_site
LOCATION: (3865)..(3866)
NAME/KEY: intron
LOCATION: (653)..(734)
US-10-066-007-4

Query Match
Best Local Similarity 100.0%; Score 3969; DB 14; Length 3969;
Matches 3969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGACCGAAGCCTCGCCAGCAGTGTGATCAAGCGAAACAGAGCGAAACAACTCTGGCGCGCC 60
DB 1 CGGACCGAAGCCTCGCCAGCAGTGTGATCAAGCGAAACAGAGCGAAACAACTCTGGCGCGCC 60
QY 61 TGGAGGAGCGGAGCGGAGGAGCAGCAGTGTGATGATCGGATGAGGAGCAGAAATCAGTAGTG 120
DB 61 TGGAGGAGCGGAGCGGAGGAGCAGCAGTGTGATGATCGGATGAGGAGCAGAAATCAGTAGTG 120
QY 121 TGTGTGTATGTGTAGTGTAGTGTGGTGTGCCATGTGCTTCTTATCATCATTT 180
DB 121 TGTGTGTATGTGTAGTGTAGTGTGGTGTGCCATGTGCTTCTTATCATCATTT 180
QY 181 TCTTTAAATCTCTACATTTGAATGTTTACCGGAACGGGCTTTGATGATACTACGGACCAC 240
DB 181 TCTTTAAATCTCTACATTTGAATGTTTACCGGAACGGGCTTTGATGATACTACGGACCAC 240
QY 241 GTTGTGTAAACAGTTCGATTTAGATTTAGATAGCGGATCCGTCGATCAGATCTCG 300
DB 241 GTTGTGTAAACAGTTCGATTTAGATTTAGATAGCGGATCCGTCGATCAGATCTCG 300
QY 301 ATCTAGAGCAGACATCTGGCTCGATCGGCTCTTCCGGAATACAGGACCGATCAGGGCA 360
DB 301 ATCTAGAGCAGACATCTGGCTCGATCGGCTCTTCCGGAATACAGGACCGATCAGGGCA 360
QY 361 GAGGAACCGGAGCGCGAAACGAGACAGACACATCATCATCAGGATGTCTTTTGT 420
DB 361 GAGGAACCGGAGCGCGAAACGAGACAGACACATCATCATCAGGATGTCTTTTGT 420
QY 421 GATCGTTTATACATCTACCTGCGATTTAAACCTTCTTCTTCTTCTTCTTCTTCTTCTT 480
DB 421 GATCGTTTATACATCTACCTGCGATTTAAACCTTCTTCTTCTTCTTCTTCTTCTTCTT 480
QY 481 GCATTTCTATCTCGTGTAAACATCGATCCGATTCGATTCGATTCGATTCGATTCGATTCG 540
DB 481 GCATTTCTATCTCGTGTAAACATCGATCCGATTCGATTCGATTCGATTCGATTCGATTCG 540
QY 541 ATCTTGTCTTGTCTACAGGTGCTTTAGGCTGGCTGCTTCTTCTTCTTCTTCTTCTTCTT 600
DB 541 ATCTTGTCTTGTCTACAGGTGCTTTAGGCTGGCTGCTTCTTCTTCTTCTTCTTCTTCTT 600
```

```
QY 601 TTCTTCAGTCTTTTACCTCGCTCCGAGCGGATCTTCACTGTATAACCTTTCAGGTTAAGAAT 660
DB 601 TTCTTCAGTCTTTTACCTCGCTCCGAGCGGATCTTCACTGTATAACCTTTCAGGTTAAGAAT 660
QY 661 TGAGCTCTGGAATCATGCTTGTGTAATCCTATAATCTCATTCATCTATCTCTTCTTCTT 720
DB 661 TGAGCTCTGGAATCATGCTTGTGTAATCCTATAATCTCATTCATCTATCTCTTCTTCTT 720
QY 721 CATCTCTCTTCAGGCGGAGATCATCAACTACTTTTACAGGCAATTTTTCAGACATCT 780
DB 721 CATCTCTCTTCAGGCGGAGATCATCAACTACTTTTACAGGCAATTTTTCAGACATCT 780
QY 781 CTCGTGAGTCTTTCATCTGCTCAGTCTGTCCTCAATCTTAAACGATCTGCTTAAACGCTT 840
DB 781 CTCGTGAGTCTTTCATCTGCTCAGTCTGTCCTCAATCTTAAACGATCTGCTTAAACGCTT 840
QY 841 TCGGACCGGTTCTTCTTCTATGTGAATCTGATCTTTGGTTTGTACGAGACACAGAG 900
DB 841 TCGGACCGGTTCTTCTTCTATGTGAATCTGATCTTTGGTTTGTACGAGACACAGAG 900
QY 901 CTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAATAACGGAAGCACCTCCGGTTTG 960
DB 901 CTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAATAACGGAAGCACCTCCGGTTTG 960
QY 961 CTGGATCGCTGGAGACACCGCTCTTGAATCTGACCGATCCGAAAGTCTTCAACAGTTTG 1020
DB 961 CTGGATCGCTGGAGACACCGCTCTTGAATCTGACCGATCCGAAAGTCTTCAACAGTTTG 1020
QY 1021 TCCATTCGAAACCTCATCTCTCTGCTGATCAATCAATCAATCAATCAATCAATCAATCAAT 1080
DB 1021 TCCATTCGAAACCTCATCTCTCTGCTGATCAATCAATCAATCAATCAATCAATCAATCAAT 1080
QY 1081 TGGACAGTGTGATGAAAGAGCCTTACGATCTATCGAAACCTTGGTATGCGCGCTCGAGTGC 1140
DB 1081 TGGACAGTGTGATGAAAGAGCCTTACGATCTATCGAAACCTTGGTATGCGCGCTCGAGTGC 1140
QY 1141 TCAGAAATCTACCGGAGATGGTGTGTTTACGCGGAGGAGTCTTTCAGTTCTCTTAT 1200
DB 1141 TCAGAAATCTACCGGAGATGGTGTGTTTACGCGGAGGAGTCTTTCAGTTCTCTTAT 1200
QY 1201 ATCACATCTAATTCACCTCGGCGGATTTGAACCTCAACATTTCTGACGAGCCTGTCACTTG 1260
DB 1201 ATCACATCTAATTCACCTCGGCGGATTTGAACCTCAACATTTCTGACGAGCCTGTCACTTG 1260
QY 1261 TTTTCACTTCAATGCTCGGTCGATCTTGTCTCATCTCATAGTGAAGTCAATAGCGAC 1320
DB 1261 TTTTCACTTCAATGCTCGGTCGATCTTGTCTCATCTCATAGTGAAGTCAATAGCGAC 1320
QY 1321 ATCAAGAGTCAATGCTCGGTCGATCTTGTCTCATCTCATAGTGAAGTCAATAGCGAC 1380
DB 1321 ATCAAGAGTCAATGCTCGGTCGATCTTGTCTCATCTCATAGTGAAGTCAATAGCGAC 1380
QY 1381 TCTTAGAAAAAGGTATGGAACCTTGTGACAAAGATGATGAGGATGCGGCTGAGAAGGATA 1440
DB 1381 TCTTAGAAAAAGGTATGGAACCTTGTGACAAAGATGATGAGGATGCGGCTGAGAAGGATA 1440
QY 1441 TGGCGCTGGGAGAGTCCGCGGTGAAAGAGGCAACAGACTCGAGACCGAAGAGTGC 1500
DB 1441 TGGCGCTGGGAGAGTCCGCGGTGAAAGAGGCAACAGACTCGAGACCGAAGAGTGC 1500
QY 1501 ATGTAAGGATTTGGTCTGAGTACCCGCTTATTCCTTCACTTGTATGAGCAGAGATAT 1560
DB 1501 ATGTAAGGATTTGGTCTGAGTACCCGCTTATTCCTTCACTTGTATGAGCAGAGATAT 1560
QY 1561 CAAGGAAGGTTTCAATGACTGACAAACACTATCTTACAGGCTGAGCTACTCTGGACGT 1620
DB 1561 CAAGGAAGGTTTCAATGACTGACAAACACTATCTTACAGGCTGAGCTACTCTGGACGT 1620
QY 1621 CATGGCTCTTCAGGTCAGTCTACTCTCTTATTAATGCTCCACATATGATGTCATGTA 1680
DB 1621 CATGGCTCTTCAGGTCAGTCTACTCTCTTATTAATGCTCCACATATGATGTCATGTA 1680
```



```
Db 3841 CTTACCCATATATATTATCCATCGTCTGTATATTTGTCTATCACGACGTCCTGTGTCGTC 3900
QY 3901 AACTCAATATTCAGCCTCTTCACTGCTTCTGTGCTCCATAGATGTGATCTTTCATGTTGT 3960
Db 3901 AACTCAATATTCAGCCTCTTCACTGCTTCTGTGCTCCATAGATGTGATCTTTCATGTTGT 3960
QY 3961 CGACTGCAG 3969
Db 3961 CGACTGCAG 3969

RESULT 2
US-10-066-007-2
; Sequence 2, Application US/10066007
; Publication No. US20030077691A1
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE
; CURRENT APPLICATION NUMBER: US/10/066,007
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US/09/518,386
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: EP 99104688.1
; PRIOR FILING DATE: 1999-03-09
; PRIOR APPLICATION NUMBER: EP 00101666.6
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (33)..(1706)
; NAME/KEY: polyA_site
; LOCATION: (1871)
; NAME/KEY: mRNA
; LOCATION: (14)..(1891)
US-10-066-007-2

Query Match 8.9%; Score 353.2; DB 14; Length 1932;
Best Local Similarity 99.2%; Pred. No. 6.6e-94;
Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3508 TCGCAGTGGTTGGCGATTTCTGTGCGCGAGATGAAGGCCCTTCTGTTGTGCACTCTCGC 3567
Db 1514 TTGCTTTGGTTGGCGATTTCTGTGCGCGAGATGAAGGCCCTTCTGTTGTGCACTCTCGC 1573
QY 3568 TCGGTCAGATTCGAGCCCATCATCTCTCATCCAGAGTACGACACATCACTTGTATCAT 3627
Db 1574 TCGGTCAGATTCGAGCCCATCATCTCTCATCCAGAGTACGACACATCACTTGTATCAT 1633
QY 3628 TTCCCGTCCTCGATCGTTGGTAGAGAGAGAGGGGTACCATCGGTTCCAGGTCAA 3687
Db 1634 TTCCCGTCCTCGATCGTTGGTAGAGAGAGAGGGGTACCATCGGTTTCAGGTCAA 1693
QY 3688 GCCGGTCGAATGAGTTGATTTCTTCATATGTTAAAGAGAAGTTCTATCTGAGAATGTGTG 3747
Db 1694 GCCGGTCGAATGAGTTGATTTCTTCATATGTTAAAGAGAAGTTCTATCTGAGAATGTGTG 1753
QY 3748 ACTAGGCAATGCTCTTCTTGTATCGATTTGTTTCTCATACCCGCGAGCGCTATGACT 3807
Db 1754 ACTAGGCAATGCTCTTCTTGTATCGATTTGTTTCTCATACCCGCGAGCGCTATGACT 1813
QY 3808 TCTACGTCGTCATCGTCGCTGGAATCTCTTCTTACCTATATATTATTCATCCG 3865
Db 1814 TCTACGTCGTCATCGTCGCTGGAATCTCTTCTTACCTATATATTATTCATCCG 1871
```

```
RESULT 3
US-10-027-632-264440
; Sequence 264440, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264440
; LENGTH: 1849
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-264440

Query Match 1.2%; Score 48; DB 13; Length 1849;
Best Local Similarity 51.4%; Pred. No. 0.013;
Matches 108; Conservative 1; Mismatches 101; Indels 0; Gaps 0;

QY 409 TGTCTTTTGTGATCGTTTTTATACATACCTACCGTTCGATCTACCTCTTCTTCTTCT 468
Db 1327 TTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1386
QY 469 CTTCGATCTTTCGATCTCTATCTCGTGAACATCGATCGATCTTTCGACCTACTTT 528
Db 1387 CTTCGATCTTTCGATCTCTATCTCGTGAACATCGATCGATCTTTCGACCTACTTT 1446
QY 529 CTCATATGTTTCATCTTGGTCTTCTCACAGGTGCTTTAGCGCTGCTGCTTCTCATGG 588
Db 1447 CTCATATGTTTCATCTTGGTCTTCTCACAGGTGCTTTAGCGCTGCTGCTTCTCATGG 1506
QY 589 GCATCCATAGGTTCTTTCAGTCTTTTACCTC 618
Db 1507 ATGATTATAGTCTCATTCGAGCCTTGAACCTC 1536

RESULT 4
US-10-027-632-264441
; Sequence 264441, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
```

[illegible]

DB 1507 ATGATTATAGCTCATTCGACGCTTGAATC 1536

```

; PRIOR FILLING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILLING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILLING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILLING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26441
; LENGTH: 1849
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-26441

```

DB 1507 ATGATTATAGCTCAATTCAGCCTTGAATC 1536

RESULT 7  
US-09-768-781-5  
; Sequence 5, Application US/09768781

```

; Patent No. US20020142376A1
;
; GENERAL INFORMATION:
;
; APPLICANT: MERKULOV, Gennady V. et al
;
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;
; TITLE OF INVENTION: AND USES THEREOF
;
; FILE REFERENCE: CL001057-CIP
;
; CURRENT APPLICATION NUMBER: US/09/768,781
;
; CURRENT FILING DATE: 2001-01-25
;
; NUMBER OF SEQ ID NOS: 7
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 5
;
; LENGTH: 17993
;
; TYPE: DNA
;
; ORGANISM: Human
;
; FEATURE:
;
; NAME/KEY: misc feature
;
; LOCATION: (1)..(17993)
;
; OTHER INFORMATION: n = A,T,C or G
;
; US-09-768-781-5

```



```

Db      1507 ATGATTATAGCTCATTGACGCTTGAACCT 1536
||||| | ||| ||| ||| ||| |||
RESULT 10
US-10-027-632-264438
; Sequence 264438, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264438
; LENGTH: 1849
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-264438

Query Match      1.2%; Score 46.8; DB 15; Length 1849;
Best Local Similarity 51.4%; Pred. No. 0.031;
Matches 108; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY      409 TGTCTTTTGTGATCGTTTTTACATACTACCGGTGAGTTCTAACTTCTTTCTTCTCT 468
DB      1327 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1386
QY      469 CTTGGCATCTTTGCATTTCTCTATCTCGTGTAACATCGATTCGAGTCTTGCCACCTT 528
DB      1387 CTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1446
QY      529 CTCCTATGTTTCATCTTGGTCTCTGCTCACAGGTCCTTAGGCTCGCTTCTCATGG 588
DB      1447 CTCTTTTCTTTCTTTCTTGACAGAGTCTCACTCTGTGCCAGGCTGAGTGCAGTGGC 1506
QY      589 GCATCCATAGCGTCTTCTCAAGTCTTTACCTC 618
DB      1507 ATGATTATAGCTCATTGACGCTTGAACCT 1536

```

```

RESULT 11
US-10-027-632-364439
; Sequence 264439, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIORITY FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20

```

```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264439
; LENGTH: 1849
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-264439

Query Match          1.2%; Score 46.8; DB 15; Length 1849;
Best Local Similarity 51.4%; Pred. No. 0.031;
Matches 108; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy      409  TGTCTTTTGTGATCGTTTTTACATACACCGTCGATCTAACCTCTTTCTCTTCT 468
Db      1327  TTCTTTCTCTTCTTTCTTTCTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTT 1386

Qy      469  CTTGCCATCTTTGCATTCCTATCTCGTTPAACATCGATCCGATCTTTGCGACCTACTTT 528
Db      1387  CTTCCTTCTCTTCTTTCTTCTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1446

Qy      529  CTCCATATGTTTCATCTTGTCCTGCTGCACAGGTCCTTAGGCCCTGGCTCTTTCTCATGG 588
Db      1447  CTCTTTTCTTCTTCTTGACAGAGGTCTCTACTCTGTTGCCAGGCTGGAGTGCAGTGCC 1506

Qy      589  GCATCCATAGCGTCTTCTAGTCTTTACCTC 618
Db      1507  ATGATTATAGCTCATTCGACGCTTTGAATC 1536

```

RESULT 12  
US-10-450-826-68  
; Sequence 68, Application US/10450826  
; Publication No US20040101818A1  
; GENERAL INFORMATION:  
; APPLICANT: Ji, Darren  
; APPLICANT: Axelrod, Douglas W.  
; APPLICANT: Cook, Jonathan S.  
; APPLICANT: Jaiswal, Neelam  
; APPLICANT: Eistein, Richard  
; APPLICANT: Houghton, Adam  
; APPLICANT: Merte, Lawrence  
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation  
; FILE REFERENCE: 044921-5039-WO  
; CURRENT APPLICATION NUMBER: US/10/450,826  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: US 60/255,882  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: US 60/285,691  
; PRIOR FILING DATE: 2001-04-24  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 68  
; LENGTH: 73465  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. AL109748  
US-10-450-826-68

Query Match	1.2%;	Score 46.8;	DB 17;	Length 73465;
Best Local Similarity	51.4%;	Pred. No. 0.32;		
Matches 108; Conservative		0; Mismatches 102;	Indels 0;	Gaps 0;

	Query Match	1.1%;	Score 44;	DB 13;	Length 886;
	Best Local Similarity	49.0%;	Pred. No. 0.13;		
	Matches 145;	Conservative 0;	Mismatches 150;	Indels 1;	Gaps 1;
Qy	1566	AAAGGTTCAATTGACTGACAAACACTACTCTTACCAGGGTCGAGCTACTCTGGACGTCATG	1625		
Db	581	AAAGTTTGAGTTCTGACGAAAGGCTACTCTTAAGGATACAGTAGAAGCCTGTGCGCTGC	640		
Qy	1626	CTCTTGCAGGTCAGCTCTACTCTCTCTTATAAATGCTCCACATATGATGATCATCTACTGAC	1685		
Db	641	CTTCTTCTCTTATTGCGAGCTTTCCTCAGTATCAGAAATACACATTAACCAACCAACTTGC	700		
Qy	1686	ATGCTCTTCCTATATTCGATACGACGTCATATGTCAG-G-GAATTGACTATAAGACGCACT	1744		
Db	701	TTGTCTGCTATATATGAAGCATAGGTTTCTTTCTTGGAGCTTGGGCTATGAAGCATAT	760		
Qy	1745	CGCTCCAGAACAGACCAATAGAGCTCTATGTCGCTTTTGTGCGACITACCGATGGGGTTTG	1804		
Db	761	TGSGTCTTTGGGATCAACCAAGAGGTTTACTTACACTGTCAAAGCCTTAACGNTGGGTCC	820		
Qy	1805	CTCCTACCTTGGACTCGTTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCGGA	1860		
Db	821	TTTATATTTTTTTTCTTTTGGAGTTTATTTTTTCGGAITTTCTTTTCTTTATTTCCCA	876		

```

RESULT 15
; US-10-027-632-168157
; Sequence 168157, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

```

Search completed: December 14, 2004, 15:23:40  
Job time : 1938.06 secs





Db 241 GTTGTGTAACAGTTTCGATTGAGATTACGATTAGATAGCCGATCCGTCGATCAGATCTCG 300  
Qy 301 ATCTAGAGCAGACTCTGGCTCGATCGTCTCTGCGGAAATCAGGCGACCGATCAGGGCA 360  
Db 301 ATCTAGAGCAGACTCTGGCTCGATCGTCTCTGCGGAAATCAGGCGACCGATCAGGGCA 360  
Qy 361 GAGGAAACCGCCGAGCGGAAACGAGACAGACACACATCATCATCAGCCATCTCTTTTGT 420  
Db 361 GAGGAAACCGCCGAGCGGAAACGAGACAGACACACATCATCATCAGCCATCTCTTTTGT 420  
Qy 421 GATGCTTTTACATACATACCCGTCGATCTAACCTCTCTTCTCTCTCTCTCTCTCTCT 480  
Db 421 GATGCTTTTACATACATACCCGTCGATCTAACCTCTCTTCTCTCTCTCTCTCTCTCT 480  
Qy 481 GCATCTCTATCTCGTGTAACATCGATCCGATCTCTGCGACCTACTCTTCTCCATATGTC 540  
Db 481 GCATCTCTATCTCGTGTAACATCGATCCGATCTCTGCGACCTACTCTTCTCCATATGTC 540  
Qy 541 ATCTTGCTCTGCTCAGAGTCTTTAGGCTCGCTGCTTTCTCATGCGGATCCATAGCG 600  
Db 541 ATCTTGCTCTGCTCAGAGTCTTTAGGCTCGCTGCTTTCTCATGCGGATCCATAGCG 600  
Qy 601 TTCTTCAGTCTTTACCTCGCTCGGCGGATCTTCACTGTATAACCTTCAGGCTAAGAT 660  
Db 601 TTCTTCAGTCTTTACCTCGCTCGGCGGATCTTCACTGTATAACCTTCAGGCTAAGAT 660  
Qy 661 TGAGCTCTGGAATCATGCTTGTGTAATCTTAAATCTCATCTCATCTCATCTCTCTCTCT 720  
Db 661 TGAGCTCTGGAATCATGCTTGTGTAATCTTAAATCTCATCTCATCTCATCTCTCTCTCT 720  
Qy 721 CATCTCTCTTCAGGCGGATCATACCACTACTTTACGCAATTTTACGATCTCT 780  
Db 721 CATCTCTCTTCAGGCGGATCATACCACTACTTTACGCAATTTTACGATCTCT 780  
Qy 781 CTGCTGAGTTTTCATCATCTGCTCAGTCTGCTCAATCTTACGATCTCTGCTTACGAGC 840  
Db 781 CTGCTGAGTTTTCATCATCTGCTCAGTCTGCTCAATCTTACGATCTCTGCTTACGAGC 840  
Qy 841 TCGAGCGGTTCTTCTTCTATGTGAAATCTGATCTTTGTTGTTGTTACGAGCAGAG 900  
Db 841 TCGAGCGGTTCTTCTTCTATGTGAAATCTGATCTTTGTTGTTGTTACGAGCAGAG 900  
Qy 901 CTGCTACAGGTGAGAGCAGTGGAGTACAGAGAAATACGGAAGCAGCCCTCCGCTTTG 960  
Db 901 CTGCTACAGGTGAGAGCAGTGGAGTACAGAGAAATACGGAAGCAGCCCTCCGCTTTG 960  
Qy 961 CTGGGATCGCTGGAGCACCGCTCTTGAACTCGACCGATCCGAAAGCTTCAACAGTTTG 1020  
Db 961 CTGGGATCGCTGGAGCACCGCTCTTGAACTCGACCGATCCGAAAGCTTCTCAACAGTTTG 1020  
Qy 1021 TCCATCCGAACTCTCTCTCTCTGATCAATCACTGATTTAAGCGACTTTGAA 1080  
Db 1021 TCCATCCGAACTCTCTCTCTCTGATCAATCACTGATTTAAGCGACTTTGAA 1080  
Qy 1081 TGGACAGTGTGATGAAAGAGCCCTACGACTATCCGAAACCTGATGCGCGCTCGAGTGC 1140  
Db 1081 TGGACAGTGTGATGAAAGAGCCCTACGACTATCCGAAACCTGATGCGCGCTCGAGTGC 1140  
Qy 1141 TCAGAAATGCTACCGGAGATGTTGTTGAGCGGAGGCTTTTCAAGTTCTCTTAT 1200  
Db 1141 TCAGAAATGCTACCGGAGATGTTGTTGAGCGGAGGCTTTTCAAGTTCTCTTAT 1200  
Qy 1201 ATCAGATCTAATCCACTCGGCGGATTTGAACTCAACTCAACTTCTGACGAGCCTGTCACTTG 1260  
Db 1201 ATCAGATCTAATCCACTCGGCGGATTTGAACTCAACTCAACTTCTGACGAGCCTGTCACTTG 1260  
Qy 1261 TTTTCATCTCANGTCTCGGTCGATCTTGTCTCATCTCATAGGTGAAGCTCATAAGCGAC 1320  
Db 1261 TTTTCATCTCANGTCTCGGTCGATCTTGTCTCATCTCATAGGTGAAGCTCATAAGCGAC 1320  
Qy 1321 ATCAGAGGATCATGATCCCTCTCTGTCGCTCAGGCGGTTAAGTCAAGTGGTCCCAATTT 1380  
Db 1321 ATCAGAGGATCATGATCCCTCTCTGTCGCTCAGGCGGTTAAGTCAAGTGGTCCCAATTT 1380

Qy 1381 TCTTAGAAAAAGGTATGAACTTTGTGCGAAGATGATGGAGGATCGGCTGAGAAAGATA 1440  
Db 1381 TCTTAGAAAAAGGTATGAACTTTGTGCGAAGATGATGGAGGATCGGCTGAGAAAGATA 1440  
Qy 1441 TGGCCGTGGAGAGTCCGCCGTGAAAGAGCAACACGACTCGAGACCGAGGAGTGC 1500  
Db 1441 TGGCCGTGGAGAGTCCGCCGTGAAAGAGCAACACGACTCGAGACCGAGGAGTGC 1500  
Qy 1501 ATGTAAAGGATTTGGTCTGAGTACCGCCCTATTCTTCACTTTGATGGACGAAGCATAT 1560  
Db 1501 ATGTAAAGGATTTGGTCTGAGTACCGCCCTATTCTTCACTTTGATGGACGAAGCATAT 1560  
Qy 1561 CAAGAAAGGTTTCACTGACTGACAAACATCTTACCAGGTCGAGCTTCTGACCT 1620  
Db 1561 CAAGAAAGGTTTCACTGACTGACAAACATCTTACCAGGTCGAGCTTCTGACCT 1620  
Qy 1621 CATGCTCTTGCAGTCTACTCTCTTATAAATGCTCCACATATGTATGCAATGTA 1680  
Db 1621 CATGCTCTTGCAGTCTACTCTCTTATAAATGCTCCACATATGTATGCAATGTA 1680  
Qy 1681 CTGACATGCTCTTCTATATTCGATACGAGTCATATGTCAGGATTTGACTATAGAGC 1740  
Db 1681 CTGACATGCTCTTCTATATTCGATACGAGTCATATGTCAGGATTTGACTATAGAGC 1740  
Qy 1741 GACTCGCTCCAGAAACAGCAATGAGTCTTATGCTGCTTTTGTGCGACTTACCGATGGG 1800  
Db 1741 GACTCGCTCCAGAAACAGCAATGAGTCTTATGCTGCTTTTGTGCGACTTACCGATGGG 1800  
Qy 1801 TTTGCTCTTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTGACTTCTTCCGA 1860  
Db 1801 TTTGCTCTTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTGACTTCTTCCGA 1860  
Qy 1861 ACTATGTAATGTCGCCATCTTTGATATCCAAAGATTTAGGATGTTACTTGTCTATA 1920  
Db 1861 ACTATGTAATGTCGCCATCTTTGATATCCAAAGATTTAGGATGTTACTTGTCTATA 1920  
Qy 1921 TTTCACTTATCGTGAACAGAAACCGAGACATGAGATACCTTTGACTCAAGGATTTAGCAGT 1980  
Db 1921 TTTCACTTATCGTGAACAGAAACCGAGACATGAGATACCTTTGACTCAAGGATTTAGCAGT 1980  
Qy 1981 TTTCCGAGAGTGGGATCGTAAGTCGAGATCAAGCTCTCTGAATATTTCTTGGTCTATC 2040  
Db 1981 TTTCCGAGAGTGGGATCGTAAGTCGAGATCAAGCTCTCTGAATATTTCTTGGTCTATC 2040  
Qy 2041 ATCTTAACCTCTAGGCTCATTCATCCATGTCGCAATAGGAGCTTATGGAGCAAAAGA 2100  
Db 2041 ATCTTAACCTCTAGGCTCATTCATCCATGTCGCAATAGGAGCTTATGGAGCAAAAGA 2100  
Qy 2101 AGCAGGCGCTGTTGGCTCAGCTTCCGATCAGGCTTGTGATAAAGAGATGTTCAAGGTC 2160  
Db 2101 AGCAGGCGCTGTTGGCTCAGCTTCCGATCAGGCTTGTGATAAAGAGATGTTCAAGGTC 2160  
Qy 2161 GGGATATCTAAGTCTCTAGGTTAGTAAAGCTTTTAAACGTATATACAGAGCGCGACA 2220  
Db 2161 GGGATATCTAAGTCTCTAGGTTAGTAAAGCTTTTAAACGTATATACAGAGCGCGACA 2220  
Qy 2221 TTTCTTCCCTGACAACTGTCAACATGCTTACTAGTGAAGCAAAACATCGCGCCCAAC 2280  
Db 2221 TTTCTTCCCTGACAACTGTCAACATGCTTACTAGTGAAGCAAAACATCGCGCCCAAC 2280  
Qy 2281 CTGCTGAACTCTAAAAGCTGTCGATGAGGAGTCTCGCTCAGATCAGTAACTGTTA 2340  
Db 2281 CTGCTGAACTCTAAAAGCTGTCGATGAGGAGTCTCGCTCAGATCAGTAACTGTTA 2340  
Qy 2341 TTTGCTGATATGAGTGTGATCTTTTCCCTCTCTATCTTCTAGCTGATTTAAAGCACTA 2400  
Db 2341 TTTGCTGATATGAGTGTGATCTTTTCCCTCTCTATCTTCTAGCTGATTTAAAGCACTA 2400  
Qy 2401 ATAGAGGCTTTTATGTTTCTTGTGATCAGAACTTCTTCGACAGCTTTCGACATGGATGT 2460  
Db 2401 ATAGAGGCTTTTATGTTTCTTGTGATCAGAACTTCTTCGACAGCTTTCGACATGGATGT 2460

QY 2461 TTCAACGACTCTCAGAGACAAAGCCGTTTCAGGATAAACTTCGAGAGAAAATTTGTGACA 2520  
DB 2461 TTCAACGACTCTCAGAGACAAAGCCGTTTCAGGATAAACTTCGAGAGAAAATTTGTGACA 2520  
QY 2521 TCGACACGATATGCTTACGCTGTGAGGATGTTTGTGATGCTAAATTAATCTCTCTTGCA 2580  
DB 2521 TCGACACGATATGCTTACGCTGTGAGGATGTTTGTGATGCTAAATTAATCTCTCTTGCA 2580  
QY 2581 AATGACTAAACCGGCTTCCATCTCTGATCCATTTTAGAGACGAACTTAATGCGTTGCGT 2640  
DB 2581 AATGACTAAACCGGCTTCCATCTCTGATCCATTTTAGAGACGAACTTAATGCGTTGCGT 2640  
QY 2641 TATCTCGAAGCGGTGTGTTCTCGATCTTGGTCTGTGCTTCTTCCAAATACAAATACGATTA 2700  
DB 2641 TATCTCGAAGCGGTGTGTTCTCGATCTTGGTCTGTGCTTCTTCCAAATACAAATACGATTA 2700  
QY 2701 TGCTCATCTGATTTGGCTGTACGGCTGTGGAATTTAACTAGTTTGTAAAGGAGTCTCTT 2760  
DB 2701 TGCTCATCTGATTTGGCTGTACGGCTGTGGAATTTAACTAGTTTGTAAAGGAGTCTCTT 2760  
QY 2761 CGTCTAGACCTCTAGTCCGTATGCTAACCGTGAATGCTTAAAGGATGAAGACGATATG 2820  
DB 2761 CGTCTAGACCTCTAGTCCGTATGCTAACCGTGAATGCTTAAAGGATGAAGACGATATG 2820  
QY 2821 TGCGTTTCATCAGCATATTTTCAATTCATATTCCTTTGTACATACGATACAGGCTGAC 2880  
DB 2821 TGCGTTTCATCAGCATATTTTCAATTCATATTCCTTTGTACATACGATACAGGCTGAC 2880  
QY 2881 CGAGCTCAAAATTCGGGCTCTCTCTGTGCTCTCTTTCTGGGCTTCTTATCTTCAAT 2940  
DB 2881 CGAGCTCAAAATTCGGGCTCTCTCTGTGCTCTCTTTCTGGGCTTCTTATCTTCAAT 2940  
QY 2941 CTTCAACAAAATTTGTACAGTTTCATCCACACTTGGGAGCCGTGATGCTGCGATG 3000  
DB 2941 CTTCAACAAAATTTGTACAGTTTCATCCACACTTGGGAGCCGTGATGCTGCGATG 3000  
QY 3001 GTCCGTATCAACGAGTCCGGATCACGAAAGAAACGATGCTATGCTTCGTAAGTTTC 3060  
DB 3001 GTCCGTATCAACGAGTCCGGATCACGAAAGAAACGATGCTATGCTTCGTAAGTTTC 3060  
QY 3061 CTTTATTTCATCTCGTCCATGAATAGTTTCTGATAGACCGGACCAATTCACGGTTGT 3120  
DB 3061 CTTTATTTCATCTCGTCCATGAATAGTTTCTGATAGACCGGACCAATTCACGGTTGT 3120  
QY 3121 CAACATCAATCGTTCAAAGTTTCAATTTATGAGAGATGCAAGAAATTCAGGTACAAATC 3180  
DB 3121 CAACATCAATCGTTCAAAGTTTCAATTTATGAGAGATGCAAGAAATTCAGGTACAAATC 3180  
QY 3181 GTTTTCTTTTAAAGCCAAATCGGTTTCGTATCGTAATGACCGGCTCTCTTTTAAATTC 3240  
DB 3181 GTTTTCTTTTAAAGCCAAATCGGTTTCGTATCGTAATGACCGGCTCTCTTTTAAATTC 3240  
QY 3241 TCGAAGACCGGAGAGTGGCTTGAGGACGTAACAGACTCGCTCAACAGTATTTGAAGCAC 3300  
DB 3241 TCGAAGACCGGAGAGTGGCTTGAGGACGTAACAGACTCGCTCAACAGTATTTGAAGCAC 3300  
QY 3301 CCTATGACACCGAGCGAGTGTATGTTTATGATTTATCTTTGTAATTTGCAAAA 3360  
DB 3301 CCTATGACACCGAGCGAGTGTATGTTTATGATTTATCTTTGTAATTTGCAAAA 3360  
QY 3361 CGTTGAATCTCGCGCTTCCCTTGTGTTGTTGTAATTCACAGTTATCTCTGAGACCGAGCTTG 3420  
DB 3361 CGTTGAATCTCGCGCTTCCCTTGTGTTGTTGTAATTCACAGTTATCTCTGAGACCGAGCTTG 3420  
QY 3421 CTTCTGAAGTTTCTTCTCATCTGGGCTTAGCAGTATCCGATCAGCCATCTAGTCTTTTG 3480  
DB 3421 CTTCTGAAGTTTCTTCTCATCTGGGCTTAGCAGTATCCGATCAGCCATCTAGTCTTTTG 3480  
QY 3481 TACGATTTGTTCTGACTCTCTCGACTTTCGAGTTTCGAGTGTGCGATTTGCTGTCGCCAGAT 3540  
DB 3481 TACGATTTGTTCTGACTCTCTCGACTTTCGAGTGTGCGATTTGCTGTCGCCAGAT 3540  
QY 3541 GAAGGCCCTTCTTGTGTTGCTACTCTCCGTCGGGTCAGTTTCGAGCCCATCTCTCATCC 3600

DB 3541 GAAGGCCCTTCTTGTGTTGCTACTCTCCGTCGGGTCAGTTTCGAGCCCATCTCTCTCATCC 3600  
QY 3601 AGAGTACGAGCACATCACCTTTCGATCATTTCCGTCCTCGAATCGTTGTTAGAGAGAAGCA 3660  
DB 3601 AGAGTACGAGCACATCACCTTTCGATCATTTCCGTCCTCGAATCGTTGTTAGAGAGAAGCA 3660  
QY 3661 GGGGTACAGATGCGGTTTTCGAGGTCAAGCCGCTCGAATAGTTGATTTTCATATGTTAA 3720  
DB 3661 GGGGTACAGATGCGGTTTTCGAGGTCAAGCCGCTCGAATAGTTGATTTTCATATGTTAA 3720  
QY 3721 GAGAAGTTCTATATCTAGAATGTTGCTAGGACAAATGCTTCTTTGATCCGATTTGTT 3780  
DB 3721 GAGAAGTTCTATATCTAGAATGTTGCTAGGACAAATGCTTCTTTGATCCGATTTGTT 3780  
QY 3781 TCTCATATCCCGGACGCGCTATGACTTCTACGTCGTCATATCGTCGCTCTCGACTCTCT 3840  
DB 3781 TCTCATATCCCGGACGCGCTATGACTTCTACGTCGTCATATCGTCGCTCTCGACTCTCT 3840  
QY 3841 CTTACCTATATATATTTCCATCGTCTGTATATTTGCTATATCAGGACGCTGTGTGCTC 3900  
DB 3841 CTTACCTATATATATTTCCATCGTCTGTATATTTGCTATATCAGGACGCTGTGTGCTC 3900  
QY 3901 AACTCAATATTCAGGCTCTTTCATGCTTCTGCTGCTCCATAGATGTGATCTTCATGTTGT 3960  
DB 3901 AACTCAATATTCAGGCTCTTTCATGCTTCTGCTGCTCCATAGATGTGATCTTCATGTTGT 3960  
QY 3961 CGACTGCGAG 3969  
DB 3961 CGACTGCGAG 3969

## RESULT 2

AX034666  
LOCUS 3969 bp DNA linear PAT 15-NOV-2000  
DEFINITION Sequence 3 from Patent EP1035206.  
ACCESSION AX034666  
VERSION AX034666.1 GI:11190640  
KEYWORDS  
SOURCE Xanthophyllomyces dendrorhous (anamorph: Phaffia rhodozyma)  
ORGANISM Xanthophyllomyces dendrorhous  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes;  
Heterobasidiomycetes; Tremellomycetidae; Cystofilobasidiales;  
Cystofilobasidiaceae; Xanthophyllomyces.

## REFERENCE

1  
Hoshino, T., Ojima, K. and Setoguchi, Y.  
Asxanthin synthetase  
Patent: EP 1035206-A 3 13-SEP-2000;  
HOFFMANN LA ROCHE (CH)

## JOURNAL

Location/Qualifiers  
1. 3969

## FEATURES

source  
/organism="Xanthophyllomyces dendrorhous"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5421"

## 5' UTR

517..518  
/note="EXPERIMENTAL"

## exon

535..652

## intron

653..734

## exon

735..783

## intron

784..898

## exon

899..1015

## intron

1016..1087

## exon

1088..1179

## intron

1180..1302

## exon

1303..1517

## intron

1518..1600

## exon

1601..1634

## intron

1635..1723

## exon

1724..1866

## intron

1867..1939

## exon

1940..1999

## intron

2000..2081

## exon

2082..2181

intron	2182..2257	
exon	2258..2354	
intron	2355..2431	
exon	2432..2542	
intron	2543..2618	
exon	2619..2652	
intron	2653..2742	
exon	2743..2814	
intron	2815..2962	
exon	2963..3050	
intron	3051..3113	
exon	3114..3171	
intron	3172..3247	
exon	3248..3321	
intron	3322..3398	
exon	3399..3423	
intron	3424..3513	
exon	3514..3700	
polyA_site	3885..3886	
/note="EXPERIMENTAL"		
ORIGIN		
Query Match 100.0%; Score 3969; DB 6; Length 3969;		
Best Local Similarity 100.0%; Pred No. 0;		
Matches 3969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	CGGACCGAAGCCTCGCCAGCAGTTGATCAAGCGAACCAAGCCGAAACAAATCCTGGGCGCC 60
Db	1	CGGACCGAAGCCTCGCCAGCAGTTGATCAAGCGAACCAAGCCGAAACAAATCCTGGGCGCC 60
Qy	61	TGGAGGCGGGAGCGGAGAGCAGCAGGTGATCGGATGATCGGTTGACAGAAACAGTATG 120
Db	61	TGGAGGCGGGAGCGGAGAGCAGCAGGTGATCGGATGATCGGTTGACAGAAACAGTATG 120
Qy	121	TGTGTGATGTGTAGTGTAGTTGGGTTGTCCTCATGTGCTTCTTCTTATCATCATCAT 180
Db	121	TGTGTGATGTGTAGTGTAGTTGGGTTGTCCTCATGTGCTTCTTCTTATCATCATCAT 180
Qy	181	TCCTTAAATCTCTACATTTGAATTTTACCGGAACGGGCTTTCATGATGATGATGATG 240
Db	181	TCCTTAAATCTCTACATTTGAATTTTACCGGAACGGGCTTTCATGATGATGATGATG 240
Qy	241	GTGTGTAAACAGTTCCGATTTGAGTTACGATTAGATGAGCGATCCCGTCCGATCAG 300
Db	241	GTGTGTAAACAGTTCCGATTTGAGTTACGATTAGATGAGCGATCCCGTCCGATCAG 300
Qy	301	ATCTAGAGCAGATCTGGCTCGATTCGGTCTTCGCGAAATCAGGCGACCGATCAGGGCA 360
Db	301	ATCTAGAGCAGATCTGGCTCGATTCGGTCTTCGCGAAATCAGGCGACCGATCAGGGCA 360
Qy	361	GAGGACCGCGAGCGCGGAGCAGACAGACACACATCATCATCAGCCATGTCTTTTGT 420
Db	361	GAGGACCGCGAGCGCGGAGCAGACAGACACACATCATCATCAGCCATGTCTTTTGT 420
Qy	421	GATCGTTTTTACATPACTACCCGTCGATTCGATTCGATTCGATTCGATTCGATTCG 480
Db	421	GATCGTTTTTACATPACTACCCGTCGATTCGATTCGATTCGATTCGATTCGATTCG 480
Qy	481	GCATTCCTATCTCGGTACATCATCGATTCGATTCGATTCGATTCGATTCGATTCG 540
Db	481	GCATTCCTATCTCGGTACATCATCGATTCGATTCGATTCGATTCGATTCGATTCG 540
Qy	541	ATCTTGTCTTGTCTACAGGTGCTTTAGGCGCTGGCTGCTTCTTCTTCTTCTTCTT 600
Db	541	ATCTTGTCTTGTCTACAGGTGCTTTAGGCGCTGGCTGCTTCTTCTTCTTCTTCTT 600
Qy	601	TTCTTCTAGTTTACCTTCGCTCGAGCGGATCTTCTGATTAACCTTCAGGGTAAGAT 660
Db	601	TTCTTCTAGTTTACCTTCGCTCGAGCGGATCTTCTGATTAACCTTCAGGGTAAGAT 660
Qy	661	TGAGCTCTGAAATCATGCTTGTGTAATCCTATAATCTCATTCATTCCTTCTTCT 720
Db	661	TGAGCTCTGAAATCATGCTTGTGTAATCCTATAATCTCATTCATTCCTTCTTCT 720

Qy	721	CATCTCTCTTTCAGGCCGGAATCATCACTACTTTTACAGGCAATTTTTTACATCCT 780
Db	721	CATCTCTCTTTCAGGCCGGAATCATCACTACTTTTACAGGCAATTTTTTACATCCT 780
Qy	781	CTCGTAGTTTTCATCATTTGGCTCAGTCTGTCCTAAATCTTAAAGATCATCGCTTAAGCCTT 840
Db	781	CTCGTAGTTTTCATCATTTGGCTCAGTCTGTCCTAAATCTTAAAGATCATCGCTTAAGCCTT 840
Qy	841	TCGGACCGGTTCTTCTTCTATGTAATCTGATCTTTTGTGTTGTTTACGAGACACAGAG 900
Db	841	TCGGACCGGTTCTTCTTCTATGTAATCTGATCTTTTGTGTTGTTTACGAGACACAGAG 900
Qy	901	CTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAATACGGAAGCACCTCCGGTTG 960
Db	901	CTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAATACGGAAGCACCTCCGGTTG 960
Qy	961	CTGGATCGCTGGAGCACCGCTCTTGAATCGACCGATCCGAAAGTCTTCAACAGTTG 1020
Db	961	CTGGATCGCTGGAGCACCGCTCTTGAATCGACCGATCCGAAAGTCTTCAACAGTTG 1020
Qy	1021	TCCATCCGAAACCTCATCTCTCTGCTGATCAATTCATCTGTAGTTAACGCACTTTGAA 1080
Db	1021	TCCATCCGAAACCTCATCTCTCTGCTGATCAATTCATCTGTAGTTAACGCACTTTGAA 1080
Qy	1081	TGGACAGTGTGATGAAAGAGCCTACGACTATCCGAAACCTGTTATGCGCGCTCGAGTGC 1140
Db	1081	TGGACAGTGTGATGAAAGAGCCTACGACTATCCGAAACCTGTTATGCGCGCTCGAGTGC 1140
Qy	1141	TCAGAAATTTGCTTACCGGAGATGGTCTTGTACGGCGGAAGTCTTTTCAAGTTCTCTTAT 1200
Db	1141	TCAGAAATTTGCTTACCGGAGATGGTCTTGTACGGCGGAAGTCTTTTCAAGTTCTCTTAT 1200
Qy	1201	ATCACATCTAATTCACCTCGGCGGATTTGAATCTAACTTTCTGACGAGCCTGTCACTTG 1260
Db	1201	ATCACATCTAATTCACCTCGGCGGATTTGAATCTAACTTTCTGACGAGCCTGTCACTTG 1260
Qy	1261	TTTTTCACTTCATGCTCTCGGTGATCTTGTCTCATCTCATAGTGAAGCTCTAAAGGCA 1320
Db	1261	TTTTTCACTTCATGCTCTCGGTGATCTTGTCTCATCTCATAGTGAAGCTCTAAAGGCA 1320
Qy	1321	ATCGAAGGATCATGATCCCTCTCTGTCGCTCAGGCGTTAAAGTCAAGTCCCAATTT 1380
Db	1321	ATCGAAGGATCATGATCCCTCTCTGTCGCTCAGGCGTTAAAGTCAAGTCCCAATTT 1380
Qy	1381	TCCTTAGAAAAGGATCGAACTTGTGACAAAGATGATGAGGATCGGCTGAGAGGATA 1440
Db	1381	TCCTTAGAAAAGGATCGAACTTGTGACAAAGATGATGAGGATCGGCTGAGAGGATA 1440
Qy	1441	TGGCCGTGGGAGAGTCCGCGCGTGAAGAAAGGCAACAGACTCGAGACCGAAGGAGTGC 1500
Db	1441	TGGCCGTGGGAGAGTCCGCGCGTGAAGAAAGGCAACAGACTCGAGACCGAAGGAGTGC 1500
Qy	1501	ATGTAAAGGATTTGGTCTGAGTACCGGCTTCTTCTTACCTTGTAGGAGAGATAT 1560
Db	1501	ATGTAAAGGATTTGGTCTGAGTACCGGCTTCTTCTTACCTTGTAGGAGAGATAT 1560
Qy	1561	CAAGAAAGGTTTCAATTCAGCTGACAAACATCTTACAGGGTCGAGCTACTCTGAGACGT 1620
Db	1561	CAAGAAAGGTTTCAATTCAGCTGACAAACATCTTACAGGGTCGAGCTACTCTGAGACGT 1620
Qy	1621	CATGGCTCTTTCAGGTCAGTCTACTCTCTTATAAATGCTCCACATATGATGCAATGA 1680
Db	1621	CATGGCTCTTTCAGGTCAGTCTACTCTCTTATAAATGCTCCACATATGATGCAATGA 1680
Qy	1681	CTGACATGCTCTTCTTATTCGATACGAGTCAATGTCAGGATTTGACTATAAGAGC 1740
Db	1681	CTGACATGCTCTTCTTATTCGATACGAGTCAATGTCAGGATTTGACTATAAGAGC 1740
Qy	1741	GACTCGCTCCAGAACAAAGCAATGAGCTCTATGTCGCTTTTGTGCGATTCACGATGGG 1800
Db	1741	GACTCGCTCCAGAACAAAGCAATGAGCTCTATGTCGCTTTTGTGCGATTCACGATGGG 1800



QY 1801 TTTGCTCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCGGA 1860  
Db 1801 TTTGCTCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCGGA 1860  
QY 1861 ACTATGGTATGCTGTGCCATTCTTTGATATCCAAAGATTATGGATAGGTTACTTCTCTAAA 1920  
Db 1861 ACTATGGTATGCTGTGCCATTCTTTGATATCCAAAGATTATGGATAGGTTACTTCTCTAAA 1920  
QY 1921 TTTCACTATCGTGAACAGAAACGGAGACATGAGATACCTTTGACTCAAGGATTAGCAGT 1980  
Db 1921 TTTCACTATCGTGAACAGAAACGGAGACATGAGATACCTTTGACTCAAGGATTAGCAGT 1980  
QY 1981 TTCCGACGAGTTGGATCGTAAGTGGCAGATCAAGCTCTCTGAAATATCTTTGGTCAATC 2040  
Db 1981 TTCCGACGAGTTGGATCGTAAGTGGCAGATCAAGCTCTCTGAAATATCTTTGGTCAATC 2040  
QY 2041 ATCTTAACCTCCTAGGCTCAITTCATCCATGGTGGCAATAGGACCTTTATGGAGCAAAAGA 2100  
Db 2041 ATCTTAACCTCCTAGGCTCAITTCATCCATGGTGGCAATAGGACCTTTATGGAGCAAAAGA 2100  
QY 2101 AGCAGCGCTGTTGGCTACGCTTCCGATCAGGCTGTGTGATAAAAGGATGTTCAAGGTC 2160  
Db 2101 AGCAGCGCTGTTGGCTACGCTTCCGATCAGGCTGTGTGATAAAAGGATGTTCAAGGTC 2160  
QY 2161 GGGATATCCTAAGTCTCCTAGGTTAGTAAACGTTTTTAAACGTATATACAGAGCGGCGACA 2220  
Db 2161 GGGATATCCTAAGTCTCCTAGGTTAGTAAACGTTTTTAAACGTATATACAGAGCGGCGACA 2220  
QY 2221 TTTCTTCCCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAAAACATCGCGCGCAAC 2280  
Db 2221 TTTCTTCCCTGACAACTGTCAACATGCTCGTTACTAGTGAGAGCAAAACATCGCGCGCAAC 2280  
QY 2281 CTGCTGAACTCAAAAGCTGTCGATGAGGAGTACTCGCTCAGATCAGTAACTTAAAGCACTA 2340  
Db 2281 CTGCTGAACTCAAAAGCTGTCGATGAGGAGTACTCGCTCAGATCAGTAACTTAAAGCACTA 2340  
QY 2341 TTTGCTGATAGTGTGATACCTTTCCCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 2400  
Db 2341 TTTGCTGATAGTGTGATACCTTTCCCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 2400  
QY 2401 ATAGAGTCTTTATGTTTCTGTTGATCAGAACTTTTGAATGAGTGTGATGAGTGTGATG 2460  
Db 2401 ATAGAGTCTTTATGTTTCTGTTGATCAGAACTTTTGAATGAGTGTGATGAGTGTGATG 2460  
QY 2461 TTCAACGACTCTCAGAGACAAAGCCGTTCAAGATTAACCTTCAGAGAGAAATTTGTGAGA 2520  
Db 2461 TTCAACGACTCTCAGAGACAAAGCCGTTCAAGATTAACCTTCAGAGAGAAATTTGTGAGA 2520  
QY 2521 TCGACACGGATATGCCCTACGCTGTGAGATGTTTGTGATGCTAAATTTACTCTCTCTGCA 2580  
Db 2521 TCGACACGGATATGCCCTACGCTGTGAGATGTTTGTGATGCTAAATTTACTCTCTCTGCA 2580  
QY 2581 AATGACTAAACAGGCTTCCATCTTGATCCATTTTAGACGAACTTTAATGCGTTGCGCT 2640  
Db 2581 AATGACTAAACAGGCTTCCATCTTGATCCATTTTAGACGAACTTTAATGCGTTGCGCT 2640  
QY 2641 TATCTCGAAGGGTGTGTTCTCGATTCTTGCTCTGCTCTTCCAAATACAAATACGATAT 2700  
Db 2641 TATCTCGAAGGGTGTGTTCTCGATTCTTGCTCTGCTCTTCCAAATACAAATACGATAT 2700  
QY 2701 TGCTCATCTGATTTGCGCTACGGGCTGTGAAATTTAAGTATTTGTTAGGAGTCTCTT 2760  
Db 2701 TGCTCATCTGATTTGCGCTACGGGCTGTGAAATTTAAGTATTTGTTAGGAGTCTCTT 2760  
QY 2761 CGTCTAGACCTCTCTAGTCCGATGCTAAACCGTGAATGCTTTAAAGGATGAAGACGTATGT 2820  
Db 2761 CGTCTAGACCTCTCTAGTCCGATGCTAAACCGTGAATGCTTTAAAGGATGAAGACGTATGT 2820  
QY 2821 TGGCTTTCATACGATATTTTCAATTTTCAATTTTCCCTTTGTACATAGCATACAGGCTGAC 2880  
Db 2821 TGGCTTTCATACGATATTTTCAATTTTCAATTTTCCCTTTGTACATAGCATACAGGCTGAC 2880  
QY 2881 CGAGCTCAAAATTCGGCTTCTCTCTCTGCTCTCTCTTTCTGCGCTTCTCTTATCTTCAAT 2940

Db 2881 CGAGCTCAAAATTCGGCTTCTCTCTCTGCTCTCTTTTCTGCGCTTCTCTTATCTTCAAT 2940  
QY 2941 CTTCAACCAAAATTTGTCCAGTTCATCCACTTCGCGAGCTGTCAATTTGTCGAGATGG 3000  
Db 2941 CTTCAACCAAAATTTGTCCAGTTCATCCACTTCGCGAGCTGTCAATTTGTCGAGATGG 3000  
QY 3001 GTCCGTCATCAACGAGTCCGATCAAGAAAGAAACGATGCTGTCTGTAAGTTTC 3060  
Db 3001 GTCCGTCATCAACGAGTCCGATCAAGAAAGAAACGATGCTGTCTGTAAGTTTC 3060  
QY 3061 CTTTATTTTCAATCTCGTCCATGAAATAGTTTCTGATAGACGGGCAATTCAGCGTTGTT 3120  
Db 3061 CTTTATTTTCAATCTCGTCCATGAAATAGTTTCTGATAGACGGGCAATTCAGCGTTGTT 3120  
QY 3121 CAACATCAATCGTTCAAAAGTTCATTTATGGAGAAATGAGTATTCAGGTACAAATTC 3180  
Db 3121 CAACATCAATCGTTCAAAAGTTCATTTATGGAGAAATGAGTATTCAGGTACAAATTC 3180  
QY 3181 GTTTTCTTTTAAAGCAATCGTTTCTGATCGTAAATTTGACCGGCTCTCTTTTAAATTC 3240  
Db 3181 GTTTTCTTTTAAAGCAATCGTTTCTGATCGTAAATTTGACCGGCTCTCTTTTAAATTC 3240  
QY 3241 TCGAAAGACCGGAGAGTGGCTTGAGGAGCTGTAACAGACTCGCTCAACAGTATTTGAAGCAC 3300  
Db 3241 TCGAAAGACCGGAGAGTGGCTTGAGGAGCTGTAACAGACTCGCTCAACAGTATTTGAAGCAC 3300  
QY 3301 CTTATGGAACACGAGGAGCTGTATGTTTTATTTATTTATCTTTGTGAATTTTCAAAA 3360  
Db 3301 CTTATGGAACACGAGGAGCTGTATGTTTTATTTATTTATCTTTGTGAATTTTCAAAA 3360  
QY 3361 CGTTGAACCTTCGCGCTTCCCTGTTGTTGAATCCAGTATCTCTGAGCCAGAGCTTG 3420  
Db 3361 CGTTGAACCTTCGCGCTTCCCTGTTGTTGAATCCAGTATCTCTGAGCCAGAGCTTG 3420  
QY 3421 CTTTGAAGTTCTTCTCATCTGGCGCTTAGCAGTATCCGATCAGCCATCTAGTTCTTTG 3480  
Db 3421 CTTTGAAGTTCTTCTCATCTGGCGCTTAGCAGTATCCGATCAGCCATCTAGTTCTTTG 3480  
QY 3481 TAGGATGCTTCTGACTCTCTCGACTTTCGAGTGTGGGATGCTGTGCGCGAGAT 3540  
Db 3481 TAGGATGCTTCTGACTCTCTCGACTTTCGAGTGTGGGATGCTGTGCGCGAGAT 3540  
QY 3541 GAAGGCTTCTTGTGTTGTCACTCTCCGTCGAGTCCAGTTCGAGCCCATCTCTCAATCC 3600  
Db 3541 GAAGGCTTCTTGTGTTGTCACTCTCCGTCGAGTCCAGTTCGAGCCCATCTCTCAATCC 3600  
QY 3601 AGAGTACAGACATCACTTGTATCAATTTCCGCTCTCGAATCGTTGTGAGAGAGGA 3660  
Db 3601 AGAGTACAGACATCACTTGTATCAATTTCCGCTCTCGAATCGTTGTGAGAGAGGA 3660  
QY 3661 GGGGTACAGATGCGTTTTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 3720  
Db 3661 GGGGTACAGATGCGTTTTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 3720  
QY 3721 GAGAGTCTCTATCTAGAGATGTTGATAGGACAATGCTTCTTTGTATCGATTTGTT 3780  
Db 3721 GAGAGTCTCTATCTAGAGATGTTGATAGGACAATGCTTCTTTGTATCGATTTGTT 3780  
QY 3781 TCTCATACCGGCGAGGCTATGACTTCTAGCTCTATCGTCTATCGTCTGACTCTCT 3840  
Db 3781 TCTCATACCGGCGAGGCTATGACTTCTAGCTCTATCGTCTATCGTCTGACTCTCT 3840  
QY 3841 CTTACCTATATATTTTCCATCGCTCTGTATATTTGTCTATCAGCAGCTCTGTGTCGTC 3900  
Db 3841 CTTACCTATATATTTTCCATCGCTCTGTATATTTGTCTATCAGCAGCTCTGTGTCGTC 3900  
QY 3901 AACTCAATTTTCAAGCTCTTCACTGCTGTCTCTCCATAGATGTGATCTTCAATTTGTT 3960  
Db 3901 AACTCAATTTTCAAGCTCTTCACTGCTGTCTCTCCATAGATGTGATCTTCAATTTGTT 3960  
QY 3961 CGACTGCGAG 3969



```
QY 325 CGGTCTTGGCGAAATCAGGACCGGATCAGGAGGAGGAAACGCCGAGCGGACGAG 384
D 1016 CGCTCACAGAAATTAATCCGAGCTTGGCTGCGAGGTCGAGGAGCTTGCATYYYYYY 1075
QY 385 CAGACACACCATCATCAGCCATGCTTTTGTGATCGTTTACATACCTACCCGTC 444
D 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135
QY 445 GATTCAACCTTCTTCTCTCTCCATCTTTCATCTCTATCTCTGTAACATC 504
D 1136 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1195
QY 505 GATCGATCTTGCACACTACTTCTCCATATGATTCATCTTGGCTTGCACAGGTC 564
D 1196 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1255
QY 565 TTAGCGCTGGCTTCTCTCATGCGCATCATAGGTTCTTCTACGCTTACCTCGCTCG 624
D 1256 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1315
QY 625 AGCGATCTCAGTGATAACCTTCAGGTAAGAAATTCAGCTTCGGAATCATGCTTGT 684
D 1316 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375
QY 685 AAATCCTATATCATCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 744
D 1376 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1435
QY 745 TACCAACTACTT 756
D 1436 TACCAACTCTT 1447

RESULT 6
AX655393/c
LOCUS AX655393 2000 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 5263 from Patent WO03000898.
ACCESSION AX655393
VERSION AX655393.1 GI:29158207
KEYWORDS Oryza sativa
SOURCE Oryza sativa
ORGANISM Oryza sativa
REFERENCE 1
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 5263 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
    source
        1..2000
            /organism="Oryza sativa"
            /mol_type="unassigned DNA"
            /db_xref="taxon:4530"

Query Match 1.5%; Score 60; DB 6; Length 2000;
Best Local Similarity 9.7%; Pred.No.1.6e-05;
Matches 63; Conservative 30; Mismatches 286; Indels 2; Gaps 1;

QY 2372 TCTATCTTCTAGCTGATTAAAGCACTAATAGAGGCTTTATGTTCTCTGTTGATCAG 2431
D 891 TYKWSRMYTWTWKAWTWTWCMAKWTATGATGATGATGATGATGATGATGATGAT 832
QY 2432 AACTTCTTCAGAGTCTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2491
D 831 MAMTKWTTTACAWRATSWRWRMAGRWKRYKYGKAYWWRWRCWKAGNARWMSRYRW 772
QY 2492 GGATAAACTTCGAGAGAAATTTGTGATGATGATGATGATGATGATGATGATGATG 2551
D 771 KWKKYATRYVYKWMWMTWMSWRNKSRYNWSGMRNRSARWYCSRMKCAKTYASSAR 712
```

```
QY 2552 TTTTGTATGCTAAATTAATCTTCTTCTGCAATAGCTAAACAGGCTTCCATCTTCAATCC 2611
D 711 WTKRKRKYRYYRRRWKKGWTKYRYRYSRMTFRMSKRKRWAGASMKSCWMYWRGA 652
QY 2612 ATTTTAGACGACCAATTAATGCTTATCTGCAAGCGGTTGGTTCTCGATTCTTGG 2671
D 651 RSMWYSKYSKSAKCKKTRYMTSSYMSTGMWYSYKSMST--SKMSYMGKMTCTMT 594
QY 2672 TCTTGTCTTCCAAATACAATAGGATTAATGCTCATCTGATTTGGCTTACGGGCTGG 2731
D 593 SMKGSFRSRKMGKRWGSMSEMYRWWKMKRKYMYRMKWKTRRCMCYRWGYTMYTTSR 534
QY 2732 AATTAACTAGTTTGTGAAGGAGTCTCTTCTGCTAGACCTCTAGTCGATGATGATC 2791
D 533 SRMYTGRYKARYTSKRYMTKYRYCHYYYGYMVKCSYMYRIGYCKACKCCYAMCW 474
QY 2792 GTGAATGCTTAAAGGATGAAGAGCTATGTTGGCTTCAACGATCAATATTTCAATTCATA 2851
D 473 KAAYSGMWYMYRYKYSKWRMSTKYMSWYKCRSMKYGAKGYCGKWTYCSYGMKW 414
QY 2852 TTCTTTTGTACATACGATACAGGCTGACGAGCTCAATTCGGGCTTCTCTTCTGTC 2911
D 413 YTMGSYKYSRYCYMYRMWYKGMWYMYYSYSSMTWYIYKYWKYWKRGTMWY 354
QY 2912 TTCTTTTCTGCGCTTCTTATCTTCTTCAATCTTCAACCAAAATTTGTACAGTTCA 2971
D 353 GKSYKKYCTWYCMYKMCYRWRWKRKTKYKRCYCWRYATCYWCCYKRGWYSR 294
QY 2972 CTTCGGAGCTGTCTATGTCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 3023
D 293 SMRTAGKWRKMSRWSRWCRCRYSWYXZYKMKWCKSYNGYSGWARRSGTWSRSA 242

RESULT 7
CCRS12862 422 bp DNA linear MAM 07-APR-2004
LOCUS Crocuta crocuta microsatellite DNA, clone T01D06.
ACCESSION AJ512862
VERSION AJ512862.2 GI:25045800
KEYWORDS microsatellite; repetitive DNA.
SOURCE Crocuta crocuta (spotted hyena)
ORGANISM Crocuta crocuta
REFERENCE 1
AUTHORS Wilhelm,K., Dawson,D.A., Gentile,L.K., Horsfield,G.F.,
Schlotterer,C., Greig,C., East,M., Hofer,H., Tautz,D. and Burke,T.
TITLE Characterisation of spotted hyena Crocuta crocuta microsatellite
loci
JOURNAL Mol. Ecol. Notes 3 (3), 360-362 (2003)
REFERENCE 2
AUTHORS Dawson,D.A.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2002) Dawson D.A., Animal & Plant Sciences,
University of Sheffield, Alfred Denny Building, Western Bank,
Sheffield, S10 2TN, UNITED KINGDOM
REMARK revised by [3]
REFERENCE 3
AUTHORS Dawson,D.A.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Dawson D.A., Animal & Plant Sciences,
University of Sheffield, Alfred Denny Building, Western Bank,
Sheffield, S10 2TN, UNITED KINGDOM
COMMENT On Nov 15, 2002 this sequence version replaced gi:24412807.
FEATURES
    source
        1..422
            /organism="Crocuta crocuta"
            /mol_type="genomic DNA"
            /db_xref="taxon:9678"
            /clone="T01D06"
            /rpt_family="microsatellite"
            51..422
                satellite
```





PN	JP	2000060568-A/2
PD	23-FEB-2000	
PR	26-AUG-1998	JP 1998239998
PI	KOICHI MIZUNO, TOMOHIKO KATO, SHIGERU SATO, DAISUKE SHIBATA	PC
	C12N15/09,A01H5/00,C07K16/40,C12N5/10,C12N9/10,C12P21/02//	PC
	(C12N15/09,C12R1:91),(C12N5/10,C12R1:91),(C12P21/02,C12R1:91),	PC
	C12N15/00,	PC
	C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91)	CC
FH	key	Location/Qualifiers
FT	CDS	(3)..(1955).
FEATURES		
source		
ORIGIN		
Query Match 1.3%; Score 50.4; DB 6; Length 6741;		
Best Local Similarity 18.6%; Pred.No.0.016;		
Matches 130; Conservative 212; Mismatches 358; Indels 0; Gaps 0;		
QY	436	CTACCGTCGATTCTTAACCTCTTCTTCTCTTCTGTCATCTTGCATCTTCTATCTCG 495
Db	6660	YDANGSTAKARCVDATBSRTTTTNSISTYCNCTBYVRANSTCRCAKYDACYTNCNC 6601
QY	496	TGTAACATCGATCCGATTTCCGACCTACTTCTTCATATGTTTCATCTTGCTCTGCTC 555
Db	6600	YYSTTNSITTTVSYTTCYTCNVCYSYSTTSTCYTTBYAKRCRTBSTYRAYTBSYDAK 6541
QY	556	ACAGTGTCTTAGGCTGCTGCTTCTCATGGCATCCATAGCGTTCTTCAGTCTTTTAC 615
Db	6540	RRCRTTBRCYAKYTBSTYAKRCCTCYVYRACYRYRATBYCRCCYIYAKNCNCYAKA 6481
QY	616	CTCGCTCCGAGCGATCTTCACTGTATAACCTTCAGGTAAGAAATGACCTCTGGAATCA 675
Db	6480	KYTBVNCVSYNCTBTCTCYTCRCVCYTCNRCNCRNCNCTBAKTBDDNSTY 6421
QY	676	TGCTTGTAATCTTAATCTCATTCATCTTCTTCTTCTTCTTCTTCTTCTCTCTCTCAG 735
Db	6420	TBYTTTTCYCYTAKCYTBGNCNCRACRDCRTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5361
QY	736	CCGGAATCATACCACTACTTTTACAGGCAATTTTATAGACATCTCTCTCGATGTTTCAT 795
Db	5360	CRYRCRCYVYRAAKYTBCTVYRCTBYSYVYAKRCTBYCTAKNCTTNDCTTNCYTTTCY 6301
QY	796	CATTGCTCAGTCGTCATCTTAACGATCATCGCTAAAGACCTTTTCGGACCGGTTCTTC 855
Db	6300	TCYTSTCSRCTNCTTDTATBTRAYRSYSTTTBAKRRNCNSTAKCTYSTTCYIYSTTCY 5241
QY	856	TTTCTATGAAATCTGATCTTGTTGTTGTTACGAGACACAGAGCTCGTACAGGTGAAG 915
Db	5240	DAYSDTTTBDRCYRCCYVTSYTERCNSTVSTCYTAKTBSRSRGYSYDAYDARDYSDCS 6181
QY	916	AGCATCGGAAGTACAGAAAAATACGAAGCACCTCCGGTTTGTGCGGATCGCTGGAG 975
Db	6180	RSRSTYSYSTSTNSTGRTBYRANSTTTRCNCRASRSTTTTCSRKYSNCCDCTYN 6121
QY	976	CACCGCTTTGAACTCGACCGATCCGAAAGCTTTCACACAGTTTGTCCATCCGAACCTC 1035
Db	6120	CSRCNSTCCYCYSSRSRNCCTTSRRTBYRAYACSRSTSTDSRSRCNSTTTCTBTTCYT 6061
QY	1036	ATCCTCTCTGCTGATCAATCACTGTAGTTTAACGCACTTTGAATGGACAGTGTATGA 1095
Db	6060	TTTTTSTTTNSTCDNSTTBDRCDSRGSYSSRRCSTRCTCTYAKTBTBTTSYDAYDA 6001
QY	1096	AGAAGCCTACGACTATCCGAAACCTGGATGCGCGCTCG 1135
Db	6000	RCRCRDYDASRSTSTYSYDAYSSITNSTSRBTBDBNS 5961
RESULT 10		
AC021625		









ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 154028)  
Donnelly,S.  
Direct Submission  
Submitted (04-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Aug 2, 2000 this sequence version replaced gi:9588479.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP4-797K12 is from the library RPCI-4 constructed at the Roswell  
Park Cancer Institute by the group of Pierer de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP4-797K12 The true  
left end of clone RP4-732F8 is at 79453 in this sequence. The true  
left end of clone RP11-526H8 is at 74481 in this sequence. The true  
right end of clone RP11-473A10 is at 68066 in this sequence.  
Location/Qualifiers  
1..154028  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="1"  
/map="p35-36.2"  
/clone="RP4-797K12"  
/clone\_lib="RPCI-4"  
repeat\_region  
155..456  
/note="AluX repeat: matches 1..292 of consensus"  
repeat\_region  
557..618  
/note="2 copies 31 mer 95% conserved"  
repeat\_region  
1211..1684  
/note="LTR42 repeat: matches 2..495 of consensus"  
repeat\_region  
1848..2136  
/note="AluX repeat: matches 1..289 of consensus"  
repeat\_region  
2199..2415  
/note="7 copies 31 mer 69% conserved"  
repeat\_region  
2208..2401  
/note="97 copies 2 mer tt 67% conserved"  
repeat\_region  
2448..2496  
/note="AluJo repeat: matches 258..304 of consensus"  
repeat\_region  
2497..2852  
/note="THE1A repeat: matches 1..354 of consensus"  
repeat\_region  
2853..3099  
/note="AluJo repeat: matches 1..258 of consensus"  
repeat\_region  
3110..3250  
/note="MIR repeat: matches 48..192 of consensus"  
repeat\_region  
3787..3917  
/note="MIR repeat: matches 28..148 of consensus"  
repeat\_region  
4416..4473  
/note="MIR repeat: matches 77..136 of consensus"  
repeat\_region  
4805..5052



```

RESULT 14
AC121998
LOCUS
DEFINITION
AC121998
AC121998
AC121998.2 GI:23499615
VERSION
HTG.
KEYWORDS
SOURCE
MUS MUSCULUS (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 135317)
McLellan, M., Meyer, R. and Dignan, G.
The sequence of Mus musculus BAC clone RP24-318H5
Unpublished (2001)
2 (bases 1 to 135317)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 135317)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 135317)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (04-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 135317)
Wilson, R.
Direct Submission
Submitted (05-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2002 this sequence version replaced gi:21040122.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: M_BE0318H05
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org.
Location/Qualifiers
1. .135317
/organism="Mus musculus"

```

FEATURES  
source

```

repeat_region      /rpt_family="L1"
50951..51155
/rpt_family="L1"
51221..51434
/rpt_family="L1"
52832..53044
/rpt_family="MaLR"
53602..53676
/rpt_family="L1"
54511..54807
/rpt_family="B4"
55778..56656
/rpt_family="ERVK"
56657..56907
/rpt_family="ERVK"
56917..57180
/rpt_family="ERVK"
57220..57438
/rpt_family="ERVK"
57441..57542
/rpt_family="L1"
57543..57845
/rpt_family="ERVK"
57984..58346
/rpt_family="L1"
59593..59731
/rpt_family="MaLR"
59799..59917
/rpt_family="L1"
60018..60157
/rpt_family="L1"
60158..60300
/rpt_family="MaLR"
60348..61622
/rpt_family="L1"
61626..62011
/rpt_family="L1"
62012..62097
/rpt_family="MaLR"
62161..62880
/rpt_family="MER2_type"
63513..63635
/rpt_family="ERVK"
63661..63935
/rpt_family="B4"
64834..65119
/rpt_family="MaLR"
65224..65468
/rpt_family="L1"
65735..65984
/rpt_family="L1"
66078..66373
/rpt_family="B4"
74305..74440
/rpt_family="Alu"
76452..76622
/rpt_family="L1"
78010..78163

Query Match
Best Local Similarity 1.2%; Score 48.6; DB 10; Length 135317;
Matches 111; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 409 TGTCTTTTGTGATCGTTTATCATCTCGTGTAAACATCGATCCGATCTTGCACCTACTTT 528
Db 7481 TTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 7540
QY 469 CTGCGATCTTGCATCTCTATCTCGTGTAAACATCGATCCGATCTTGCACCTACTTT 528
Db 7541 CTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 7600
QY 529 CTCATATGTCATCTTGGTCTTGTCTACAGGTCCTTACAGGTCCTTGCATCTTCTCATGG 588

```

7601 CTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 7660

589 GCATCCATAGCGTTCTTTCAGTCTTTACCTCGCTCC 623

7661 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 7695

RESULT 15

AL928832 198993 bp DNA linear ROD 16-DEC-2002

LOCUS Mouse DNA sequence from clone RP23-261N18 on chromosome 2, complete sequence.

ACCESSION AL928832

VERSION AL928832.7 GI:26984867

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 198993)

AUTHORS Tracey, A.

JOURNAL Direct Submission

TITLE Submitted (14-DEC-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humqueres@sanger.ac.uk

COMMENT On Dec 15, 2002 this sequence version replaced gi:26788196. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humqueres@sanger.ac.uk](mailto:humqueres@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-261N18 is from the RPCI-23 Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

1. 198993

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="2"

/clone="RP23-261N18"

/clone\_lib="RPCI-23"

ORIGIN

Query Match 1.2%; Score 48.6; DB 10; Length 198993;

Best Local Similarity 51.6%; Pred. No. 0.12;

Matches 111; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Search completed: December 14, 2004, 09:31:08  
Job time : 16607.7 secs

Search completed: December 14, 2004, 09:31:08  
Job time : 16607.7 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model  
Run on: December 14, 2004, 01:15:27 ; Search time 5818.92 Seconds  
(without alignments)  
12098.737 Million cell updates/sec

Title: US-10-066-007A-2  
Perfect score: 1932  
Sequence: 1 gaattcgacagggccacc.....agccggctgcgcgaattc 1932

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues  
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gssi.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	4.6	528	6	CD275123
2	79.2	4.1	474	6	CD275883
3	79.2	4.1	475	6	CD275923
4	75.8	3.9	615	1	AT002896
5	69.2	3.6	354	6	CD274181
6	68	3.5	403	6	CD273194
7	61	3.2	487	5	BQ102580
8	58.8	3.0	400	6	CD275427
9	58	3.0	317	6	CD275197
10	56.8	2.9	439	4	BJ075377
11	55.4	2.9	971	5	EX779027
12	53.8	2.8	786	7	CR416865
13	53.8	2.8	849	6	CF152584
14	51.2	2.7	588	2	AW600900
15	50.6	2.6	1296	2	AW600865
16	50.4	2.6	875	6	CB200527
17	50.2	2.6	407	5	BU398544
18	50.2	2.6	723	5	BU250947
19	50.2	2.6	951	5	BU246013
20	50	2.6	508	1	AJ791608
21	50	2.6	740	4	EG207705
22	50	2.6	907	9	CNS015AK
23	49.8	2.6	212	1	AL651909
24	49.8	2.6	303	1	AA907131

25	49.6	2.6	544	7	CF386878
26	49.4	2.6	447	1	AV736322
27	49.2	2.5	387	5	BU79630
28	49	2.5	483	6	CB821156
29	49	2.5	738	7	CK694738
30	48.8	2.5	455	1	AV744471
31	48.8	2.5	730	7	CP430452
32	48.6	2.5	547	4	BI378369
33	48.4	2.5	396	4	BM139826
34	48.4	2.5	480	6	CB157627
35	48.4	2.5	580	6	CB114922
36	48.4	2.5	609	6	CB162338
37	48.4	2.5	641	6	CB154142
38	48.4	2.5	697	1	AV648984
39	48.4	2.5	755	4	EG206682
40	48.4	2.5	757	4	EG211348
41	48.4	2.5	760	4	EG196339
42	48.4	2.5	763	4	EG206681
43	48.4	2.5	765	4	EG182832
44	48.4	2.5	768	4	EG220280
45	48.4	2.5	770	4	EG197817

ALIGNMENTS

RESULT 1  
CD275123  
LOCUS  
DEFINITION  
T143B00283 (FHIG:B) Axenic plate culture Paxillus involutus cDNA  
5', mRNA sequence.  
CD275123  
VERSION  
CD275123.1 GI:34387169  
KEYWORDS  
EST.  
SOURCE  
Paxillus involutus  
ORGANISM  
Paxillus involutus  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Boletales; Paxillineae; Paxillaceae; Paxillus.  
REFERENCE  
1 (bases 1 to 528)  
AUTHORS  
Johansson, T., Le Quere, A., Ahren, D., Soderstrom, B., Erlandsson, R., Lundberg, J., Uhlen, M. and Tunlid, A.  
TITLE  
Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue  
JOURNAL  
Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
COMMENT  
Contact: Johansson, T.  
Fungal-Host Interaction Group (FHIG)  
Microbial Ecology, Institution of Ecology  
Ecology Building, Lund University, SE-223 62 Lund, Sweden  
Tel: +46 46 222 45 43  
Fax: +46 46 222 41 58  
Email: tomas.johansson@bioekol.lu.se  
PCR Primers  
FORWARD: P104 (5'-GGGAAGCGCGCATTTGTT-3')  
BACKWARD: P105 (5'-AGTGAGTCGATTCGGCC-3')  
Seq primer: P104  
High quality sequence stop: 528.  
Location/Qualifiers  
1. 528  
/organism="Paxillus involutus"  
/mol\_type="mRNA"  
/strain="ATCC 200175"  
/db\_xref="taxon:71150"  
/tissue\_type="mycelium"  
/dev\_stage="25 days of growth after transfer"  
/lab\_host="Escherichia coli BM25.8"  
/clone\_lib="(FHIG:B) Axenic plate culture"  
/note="Vector: pTriplex2; Site 1: Sfil; Site 2: Sfil; This EST clone is originating from one of three cDNA libraries, constructed for transcript profiling of the mycorrhizal interaction between the basidiomycete Paxillus involutus and Betula pendula (birch). One library represents the developed and functional mycorrhizal root tissue ('(FHIG:A) Ectomycorrhiza plate culture'), a second

FEATURES  
source

library represents axenically grown fungus ('(FHIG:B) Axenic plate culture') and a third library represents axenically grown plants ('(FHIG:C) Axenic plate culture'). Libraries were analyzed in parallel and 3555 (FHIG:A), 3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20) ESTs of ~99bp have been deposited. The cDNA libraries were constructed from total RNA using the SMART cDNA library construction kit (#K1051-1, Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Full-length cDNAs were trimmed by SfiI, fractionated and directionally ligated into (lambda) triplex2 arms. The lambda library was converted to a plasmid library via site-specific recombination at loxP sites in a Cre+ strain (E. coli BM25.8). Plasmid clones were randomly collected and analysed by DNA sequencing using a plasmid-specific forward primer (F104)."

## ORIGIN

Query Match 4.1%; Score 88; DB 6; Length 528;  
Best Local Similarity 54.6%; Pred. No. 7.7e-13;  
Matches 244; Conservative 0; Mismatches 195; Indels 9; Gaps 3;

QY 1069 CAGTCTTGACATGATGTTTCACCGACTCTCAGAGAACAAAGCCGTTTCAGGATAAATCTTC 1128  
DB 88 CAGCGATAACTTGGCGCTGCATGAGTTGTCTGCTAGCCCTGGGATACAGAGGAAGCTCC 147

QY 1129 GAGBAGAAATTTGTCAGATCGACACGGATATGCTAGCGTAGACGAATTAATGCTTGC 1188  
DB 148 GGGAGGAACACTTCTTCTGTTGACACGGAACGCCCTCGATGGAGCAACTCTCGGCGCTTC 207

QY 1189 CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTCGTCTAGACCTCTCTAGTCGATGCTA 1248  
DB 208 CTTACCTTGACACCGTGTGAAGGAGACTTTCGTGTGTACACCTCTCTTGGAGAGACCA 267

QY 1249 ACCGTGAATGTTAAAGTGTAGAGACTTCATCCCACTTCGCGAGCTGTCATGTCGAG 1308  
DB 268 TACGCG---TGGCATGAAAGAGATATCTGCGCTCGAGAGCCCACTTCAGGACAAAC 324

QY 1309 ATGGGTGGTTCATCAACGAGTTCGGATCAGAAAGGAACGATGTCATGCTTCCTCGTTGT 1368  
DB 325 ATGCGCTGTACACGACGGAATAGATCAGAAAGGACCGGATCTCTATCCGATCT 384

QY 1369 TCAACATCAATCGTTTAAAGTTTATTTATGGAAGATGAGAGAAATTCAGACCGGAGA 1428  
DB 385 T---GATGAACCGGTGCAAGGGGTGTGGGGCCCTGACGCCCAAGAGTTCAAAACCGGAAC 441

QY 1429 GTGCGCTTGAGGAGTACAGACTCGCTCAACAGTATTGAGCACCTATGGACACGAG 1488  
DB 442 GCTGG---GACGAGTACCTTAAGCTGTCTCCATATTCAGGTGTTTGGGGCCACATGC 498

QY 1489 CGAGCTTTATCTCTGGACCCAGAGCTTG 1516  
DB 499 TAAGTTTCTCGGGGGCCAGGACATG 526

## RESULT 2

CD275883  
LOCUS T143B05239F (FHIG:B) 474 bp mRNA linear EST 01-SEP-2003  
DEFINITION 5', mRNA sequence.  
ACCESSION CD275883  
VERSION CD275883.1 GI:34387929  
KEYWORDS EST.  
SOURCE Paxillus involutus  
ORGANISM Paxillus involutus  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Boletales; Paxillineae; Paxillaceae; Paxillus.  
1 (bases 1 to 474)  
REFERENCE Johansson, T., Le Quere, A., Ahren, D., Soderstrom, B., Eriandsson, R., Lundeberg, J., Uhlen, M. and Tunlid, A.  
AUTHORS Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue  
TITLE Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
JOURNAL

## COMMENT

Contact: Johansson, T.  
Fungal-Host Interaction Group (FHIG)  
Microbial Ecology, Institution of Ecology  
Ecology Building, Lund University, SE-223 62 Lund, Sweden  
Tel: +46 46 222 45 49  
Fax: +46 46 222 41 58  
Email: tomas.johansson@mbioekol.lu.se

PCR Primers  
FORWARD: P104 (5'-GGGAGCGCGCATTTGTT-3')  
BACKWARD: P105 (5'-AGTGAGCTCGAATGGGCC-3')  
Seq primer: P104  
High quality sequence stop: 474.

## FEATURES

source

1..474  
/organism="Paxillus involutus"  
/mol\_type="mRNA"  
/strain="ATCC 200175"  
/db\_xref="taxon:71150"  
/tissue\_type="Mycelium"  
/dev\_stage="25 days of growth after transfer"  
/lab\_host="Escherichia coli BM25.8"  
/clone\_lib="(FHIG:B) Axenic plate culture"  
/note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This EST clone is originating from one of three cDNA libraries, constructed for transcript profiling of the mycorrhizal interaction between the basidiomycete Paxillus involutus and Betula pendula (birch). One library represents the developed and functional mycorrhizal root tissue ('(FHIG:A) Ectomycorrhiza plate culture'), a second library represents axenically grown fungus ('(FHIG:B) Axenic plate culture') and a third library represents axenically grown plants ('(FHIG:C) Axenic plate culture'). Libraries were analyzed in parallel and 3555 (FHIG:A), 3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20) ESTs of ~99bp have been deposited. The cDNA libraries were constructed from total RNA using the SMART cDNA library construction kit (#K1051-1, Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Full-length cDNAs were trimmed by SfiI, fractionated and directionally ligated into (lambda) triplex2 arms. The lambda library was converted to a plasmid library via site-specific recombination at loxP sites in a Cre+ strain (E. coli BM25.8). Plasmid clones were randomly collected and analysed by DNA sequencing using a plasmid-specific forward primer (F104)."

## ORIGIN

Query Match 4.1%; Score 79.2; DB 6; Length 474;  
Best Local Similarity 54.6%; Pred. No. 2.1e-10;  
Matches 203; Conservative 0; Mismatches 163; Indels 6; Gaps 2;

QY 1069 CAGTCTTGACATGATGTTTCACCGACTCTCAGAGAACAAAGCCGTTTCAGGATAAATCTTC 1128  
DB 79 CAGCGATAACTTGGCGCTGCATGAGTTGTCTGCTAGCCCTGGGATACAGAGGAAGCTCC 138

QY 1129 GAGBAGAAATTTGTCAGATCGACACGGATATGCTAGCGTAGACGAATTAATGCTTGC 1188  
DB 139 GGGAGGAACACTTCTTCTGTTGACACGGAACGCCCTCGATGGAGCAACTCTCGGCGCTTC 198

QY 1189 CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTCGTCTAGACCTCTCTAGTCGATGCTA 1248  
DB 199 CTTACCTTGACACCGTGTGAAGGAGACTTTCGTGTGTACACCTCTCTTGGAGAGACCA 258

QY 1249 ACCGTGAATGTTAAAGGATGAGACTTCATCCCACTTCGCGAGCTGTCTATGCTGCGAG 1308  
DB 259 TACGCG---TGGCATGAAAGAGATATCTGCGCTCGAGAGCCCACTTCAGGACAAAC 315

QY 1309 ATGGGTGGTTCATCAACGAGTTCGGATCAGAAAGGAACGATGTCATGCTTCGTTGT 1368  
DB 316 ATGGCGTGTACACGACCGGAATTAGGATCAGAAAGGACCGCGATCTCTATCCGATCT 375

QY 1369 TCAACATCAATCGTTCAAGTTTATTTATGGAAGATGAGAGAAATTCAGACCGGAGA 1428



Db 376 T---GATGAACGGTCGAAGGGTTGTGGGCCCTGACGCCACGAGTTCAAACCGGAAC 432

QY 1429 GGTGGCTTGAGG 1440

Db 433 CTGGGACGACG 444

RESULT 3

CD275923

LOCUS

DEFINITION

CD275923

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..475

/organism="Paxillus involutus"

/mol\_type="mRNA"

/strain="ARCC 200175"

/db\_xref="taxon:71150"

/tissue\_type="Mycelium"

/dev\_stage="25 days of growth after transfer"

/lab\_host="Escherichia coli BM25.8"

/clone\_lib="(FHIG:B) Axenic plate culture"

/note="vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This EST clone is originating from one of three cDNA libraries, constructed for transcript profiling of the mycorrhizal interaction between the basidiomycete Paxillus involutus and Betula pendula (birch). One library represents the developed and functional mycorrhizal root tissue ((FHIG:A) Ectomycorrhiza plate culture), a second library represents axenically grown fungus ((FHIG:B) Axenic plate culture) and a third library represents axenically grown plants ((FHIG:C) Axenic plate culture)". Libraries were analyzed in parallel and 3555 (FHIG:A), 3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20) ESTs of >99bp have been deposited. The cDNA libraries were constructed from total RNA using the SMART cDNA library construction kit (#K1051-1, Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Full-length cDNAs were trimmed by SfiI, fractionated and directionally ligated into (lambda)Triplex2 arms. The lambda library was recombination at loxP sites in a Cre+ strain (E. coli BM25.8). Plasmid clones were randomly collected and analysed by DNA sequencing using a plasmid-specific forward primer (P104)."

4.18; Score 79.2; DB 6; Length 475;

Best Local Similarity 54.6%; Pred. No. 2.1e-10;

Matches 203; Conservative 0; Mismatches 163; Indels 6; Gaps 2;

QY 1069 CAGTCTTGACATGGATGTTTACCGACTCTCAGAAAGACAAACCGCTTACGGATAAAGTTC 1128

Db 88 CAGCGATAAATTTGGGCGCTGCAATGATGTTGCTAGCCCTCGGGATACAGAGAGAGCTCC 147

QY 1129 GAGAAGAAATTTGTTCAGATCGACACGGATATCCCTACGCTAGACGAACCTTAATGGCTTGC 1188

Db 148 GGGAGGAATCTTCTTCTGTTGACACGGAAACGCCCTCGATGACGAACTCTCGGGGCTTC 207

QY 1189 CTTATCTCGAAGCGTTTGTAAAGAGTCTTTCGTCTAGACCCCTCCTAGTCCGATGCTA 1248

Db 208 CCTACCTTCACACCGCTGGTGAAGGAGACTTTGCGTGTACACCCCTCCCTTTGGAGAGCA 267

QY 1249 ACCGTGAATGCTTAAAGGATGAAGACTTCATCCCACTTCCGAGCCTGTCTATTGTCGAG 1308

Db 268 TAGCG---TCGCGATGAAGACGATATTCCTCCGCTCGAAGACCACTCAGGACAAC 324

QY 1309 ATGGGTTCGTATCAACAGGTCGCGATCAAGAAAGGACGATGGTCTATGCTTCGGTTGT 1368

Db 325 ATGGCGTCTACACGACGGAATTAGGATCAGCAAAGGACCGGATCTCTATCCCGATCT 384

QY 1369 TCAACATCAATGTTCAAGTTCATTATGGAGAGAGTTCAGAAAGAAATTCAGACCGGAGA 1428

Db 385 T---GATGAACCGGTGGAAGGGGTTGTGGGGCCCTGACGCCACGAGTTCAAACCGGAAC 441

QY 1429 GGTGGCTTGAGG 1440

Db 442 GCTGGGACGACG 453

RESULT 4

AT002896

LOCUS

DEFINITION

AT002896

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

1..615

/organism="Pleurotus ostreatus"

/mol\_type="mRNA"

/cultivar="ASI 2029"

/db\_xref="taxon:5322"

/clone="355LM"

/dev\_stage="shaking liquid cultured mycelia"

/lab\_host="E.coli"

/clone\_lib="POS1M01"

/note="Vector: lambda Uni-ZAP XR; Site 1: EcoRI; Site 2: XhoI; average insert size:1000 bp;initial prui:5 X 10 7

URL: http://bric.postech.ac.kr/.

Location/Qualifiers

1..615

/organism="Pleurotus ostreatus"

/mol\_type="mRNA"

/cultivar="ASI 2029"

/db\_xref="taxon:5322"

/clone="355LM"

/dev\_stage="shaking liquid cultured mycelia"

/lab\_host="E.coli"

/clone\_lib="POS1M01"

/note="Vector: lambda Uni-ZAP XR; Site 1: EcoRI; Site 2: XhoI; average insert size:1000 bp;initial prui:5 X 10 7

Library information: Isolation of total RNA from the mycelia incubated in shaking liquid MCM media at 30 deg C"

ORIGIN

Query Match 3.9%; Score 75.8; DB 1; Length 615;  
Best Local Similarity 48.9%; Pred. No. 2e-09;  
Matches 234; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 1265 GGATGAGACTTCATCCCATCTGCGAGCGTGTGTCGAGATGGGTGGTGCATCA 1324  
DB 59 GGCTGATGATGTTCTCCGTTGAGCGTCCGTTGACAGATCTGAAGGTATTAATCA 118  
QY 1325 CGAGTCCGGATCAGAAAGAACGATGTCATGCTTCCGTTGTTCAACATCAATCGTTC 1384  
DB 119 ATCAGTAGCAATCAAGAAAGGACAAATGTTTCATGTTCCGATCGTTCCGATGAATAGATT 178  
QY 1385 AAAGTTCATTCGAGAGATGAGAGAAATTCAGACCGGAGAGTGGCTTGAGGACGT 1444  
DB 179 GGTATCAATATGGCGGAGGATGCTTTGAAATCAAGCCAGAACGTTGGC---AATCACC 235  
QY 1445 AACAGACTCGCTCAACAGTATTGAAGCACCCCTATGGACACCGCGAGCTTTATCTCTGG 1504  
DB 236 GCGGAGGCTCCAGCGTGTGTCAGGGATATGGAGTAACATGATGACTTTCCTTGGTGG 295  
QY 1505 ACCAGAGCTGCTTGGTGGGATGCTGTCGCGGAGATGAGGCCCTTCTGTTGT 1564  
DB 296 TCCTCGAGCATGATCGTTGGAGATTCTCTATCTAGAGATGAAGCCCTTGTGTTAC 355  
QY 1565 CACTCTCGTGGTTCAGATTCGAGGCCATCATCTCTCCAGAGTACGAGCACATCAC 1624  
DB 356 CTTGTGGTTCATTCGAGTTCGAGCTCGCTGTTCCATCAGAGGAATGGCAGATCTC 415  
QY 1625 TTGATCATTTCCCGTCTCGAATCGTGTGAGAGAAAGAGGGGTACCAGATGGCTT 1684  
DB 416 GTCCATCTGTCAGCGACCTATTCTGAAGAACGATAAAGAGCGGGAATGTGATGCCCT 475  
QY 1695 CGAGTCAAGCGGTGAGATGATGTTCTTCATATGTTAAAGAGAGTCTATATCTG 1743  
DB 476 GATCATTAACCATATCAACGGGAGTGACCTCAACATTTAACGATGGATATGTAGCG 534

RESULT 5  
LOCUS CD274181  
DEFINITION T143B02839F (FHIG:B) Axenic plate culture Paxillus involutus CDNA  
5', mRNA sequence.  
ACCESSION CD274181  
VERSION CD274181.1 GI:34386227  
KEYWORDS EST.  
SOURCE Paxillus involutus  
ORGANISM Paxillus involutus  
REFERENCE Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
AUTHORS Boletales; Paxillineae; Paxillaceae; Paxillus.  
1 (bases 1 to 354)  
TITLE Johansson, T., Le Quere, A., Ahren, D., Soderstrom, B., Erlandsson, R.,  
Lundberg, J., Uhlen, M. and Tunlid, A.  
JOURNAL Transcriptional responses of Paxillus involutus and Betula pendula  
COMMENT during formation of ectomycorrhizal root tissue  
Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
Contact: Johansson, T.  
Fungal-Host Interaction Group (FHIG)  
Microbial Ecology, Institution of Ecology  
Ecology Building, Lund University, SE-223 62 Lund, Sweden  
Tel: +46 46 222 45 49  
Fax: +46 46 222 41 58  
Email: tomas.johansson@bioekol.lu.se  
PCR Primers  
FORWARD: P104 (5'-GGGAGCGCGCCATGTTGTT-3')  
BACKWARD: P105 (5'-AGTGGCTCGAATGTCGGCC-3')  
Seq primer: P104  
High quality sequence stop: 354.  
Location/Qualifiers  
1..354  
FEATURES  
source

/organism="Paxillus involutus"  
/mol\_type="mRNA"  
/strain="ATCC 200175"  
/db\_xref="taxon:71150"  
/tissue\_type="Mycelium"  
/dev\_stage="25 days of growth after transfer"  
/lab\_host="Escherichia coli BM25.8"  
/clone\_lib="(FHIG:B) Axenic plate culture"  
/note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This  
EST clone is originating from one of three cDNA libraries,  
constructed for transcript profiling of the mycorrhizal  
interaction between the basidiomycete Paxillus involutus  
and Betula pendula (birch). One library represents the  
developed and functional mycorrhizal root tissue  
(/FHIG:A) Ectomycorrhiza plate culture', a second  
library represents axenically grown fungus (/FHIG:B)  
Axenic plate culture', and a third library represents  
axenically grown plants (/FHIG:C) Axenic plate culture'.  
Libraries were analyzed in parallel and 355 (FHIG:A),  
394 (FHIG:B), and 232 (FHIG:C) high-quality (PHRED 20)  
ESTs of >99bp have been deposited. The cDNA libraries were  
constructed from total RNA using the SMART cDNA library  
Construction kit (#1051-1, Clontech, Palo Alto, CA, USA)  
according to the manufacturer's instructions. Full-length  
cDNAs were trimmed by SfiI, fractionated and directionally  
ligated into (lambda)Triplex2 arms. The lambda library was  
converted to a plasmid library via site-specific  
recombination at loxp sites in a Cre+ strain (E. coli  
BM25.8). Plasmid clones were randomly collected and  
analysed by DNA sequencing using a plasmid-specific  
forward primer (P104)."

## ORIGIN

Query Match 3.6%; Score 69.2; DB 6; Length 354;  
Best Local Similarity 54.4%; Pred. No. 1.2e-07;  
Matches 162; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

QY 1069 CAGTCTTCACATGGATGTTTCACCGACTCTCAGAAAGCAAGCGTTCCAGGATAAATTC 1128  
DB 60 CAGCATTAACCTTGGCGCTGTCATGATGTTGTCGTAGCCCTGGGATACAGAGGAAGCTCC 119  
QY 1129 GAGAAAGAAATTTGTCAGATCGACCGGATATGCTACGCTAGACGAACTTAATCGCTGC 1188  
DB 120 GCGAGGAACTTCTTTCTGTCACGGAACGCCCTCGATGAGCAAACTCTCGCGCTTC 179  
QY 1189 CTTATCTCGAAGCGTTTCTTAAGGAGTCTCTGCTAGACCCCTCTAGTCCGTATGCTA 1248  
DB 180 CTTACCTTGACACCGTGGTGAGGAGACTTTGCGTGTACACCTCCCTTTGTAGAGACA 239  
QY 1249 ACCGTGAATGCTTAAAGATGAAGACTTCATCCCACTTGGCGAGCTGCTCATTTGGTGCAG 1308  
DB 240 TAGCGG---TCGCGATGAAGACGATATTTCTCGCGCTCGAAGAACCACTCAGCAACAAC 296  
QY 1309 ATGGTTCGCTCATCAACGAGTCCGGATCGGATCAGAAAGGAACGATGCTATGCTTCGTT 1366  
DB 297 ATGGCGCTCGTACACGAGGAATAGGATCAGCAAAAGGACCGGATCTCTATCCGAT 354

## RESULT 6

LOCUS CD273194  
DEFINITION T143B00316 (FHIG:B) Axenic plate culture Paxillus involutus CDNA  
5', mRNA sequence.  
ACCESSION CD273194  
VERSION CD273194.1 GI:34385240  
KEYWORDS EST.  
SOURCE Paxillus involutus  
ORGANISM Paxillus involutus  
REFERENCE Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
AUTHORS Boletales; Paxillineae; Paxillaceae; Paxillus.  
1 (bases 1 to 403)  
Johansson, T., Le Quere, A., Ahren, D., Soderstrom, B., Erlandsson, R.,  
Lundberg, J., Uhlen, M. and Tunlid, A.

TITLE  
JOURNAL  
COMMENT

Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue  
Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
Contact: Johansson, I.

Fungal-Host Interaction Group (FHIG)

Microbial Ecology, Institution of Ecology

Ecology Building, Lund University, SE-223 62 Lund, Sweden

Tel: +46 46 222 45 49

Fax: +46 46 222 41 58

Email: tomas.johansson@bioekol.lu.se

PCR Primers

FORWARD: P104 (5'-GGGAAGCGCGCATTTGTT-3')

BACKWARD: P105 (5'-AGTGAGTCGAATTCGGCC-3')

Seg primer: P104

High quality sequence stop: 403.

FEATURES  
source

Location/Qualifiers  
1..403  
/organism="Paxillus involutus"  
/mol\_type="mRNA"  
/strain="ATCC 200175"  
/db\_xref="taxon:71150"  
/tissue\_type="Mycelium"  
/dev\_stage="35 days of growth after transfer"  
/lab\_host="Escherichia coli BM25.8"  
/clone\_lib="(FHIG:B) Axenic plate culture"  
/note="Vector: pTriplex2; Site 1: Sfil; Site 2: Sfil; This EST clone is originating from one of three cDNA libraries, constructed for transcript profiling of the mycorrhizal interaction between the basidiomycete Paxillus involutus and Betula pendula (birch). One library represents the developed and functional mycorrhizal root tissue ('(FHIG:A) Ectomycorrhiza plate culture'), a second library represents axenically grown fungus ('(FHIG:B) Axenic plate culture') and a third library represents axenically grown plants ('(FHIG:C) Axenic plate culture'). Libraries were analyzed in parallel and 355 (FHIG:A), 394 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20) ESTs of >99bp have been deposited. The cDNA libraries were constructed from total RNA using the SMART cDNA library construction kit (#K1051-1, Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Full-length cDNAs were trimmed by Sfil, fractionated and directionally ligated into (lambda)Triplex2 arms. The lambda library was converted into a plasmid library via site-specific recombination at loxP sites in a Cre+ strain (E. coli RM25.8). Plasmid clones were randomly collected and analysed by DNA sequencing using a plasmid-specific forward primer (P104)."

ORIGIN

Query Match 3.5%; Score 68; DB 6; Length 403;  
Best Local Similarity 53.6%; Pred. No. 2.6e-07;  
Matches 165; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 1069 CAGTCCTTGACATGGATGTTTCCCGACTCTCAGAACAGAAAGCGTTCAGGATAACTTC 1128  
DB 99 CAGCGATAATTGGGCGCTGCATGAGTTGTGCTAGCCCTCGGATACAGAGAACTCC 158  
QY 1129 GAGAAGAAATTTGTCAGATCGAACGGATATGCTACGCTAGACGAATTAATGCGTTGC 1188  
DB 159 GGGAGGAAATCTTCTTGTTGACACGAAACGGCCCTCGATGGACGAATCTCGGCGCTTC 218  
QY 1189 CTTATCTCAGAGCGTTTGTAAAGAGTCCTTCTGTAGACCGCTCTAGTCGGATGCTA 1248  
DB 219 CTTACCTTGACACGTTGGGAAGAGACTTTGCTGTACACCTCTCCTTTGGAGACCA 278  
QY 1249 ACCGTGAATGCTTTAAAGGATGAAGACTTTCATCCACTTTGCCGAGCGCTGTCAITGGTCGAG 1308  
DB 279 TACGCG---TCGCGATGAAGACGATATTCTCGCGCTCGAGAGCCACTCACGGAACAAC 335  
QY 1309 ATGGTTCGTCATCAACGAGTCCGATACGAAAGACGATGTCATGCTTCGTTGT 1368  
DB 336 ATGGCGCTGTACACGACGGAATTAGGATCAGCAGAGGACCGCGATCTCTATCCCGATCT 395

QY 1369 TCAACATC 1376  
DB 396 TGATGAAC 403

RESULT 7  
BQ102580

LOCUS  
DEFINITION

BQ102580 487 bp mRNA linear EST 12-APR-2002  
MIN155 MIN Nitrogen-replete Schizophyllum library Schizophyllum commune cDNA 5' similar to cytochrome p450, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..487

/organism="Schizophyllum commune"

/mol\_type="mRNA"

/strain="4-40"

/db\_xref="taxon:5334"

/tissue\_type="mycelium"

/clone\_lib="MIN Nitrogen-replete Schizophyllum library"

/note="Vector: lambda Zap; Site 1: EcoRI; Site 2: XhoI; 4-day-old mycelia of Schizophyllum commune were transferred from minimal (nitrogen-replete) medium to fresh minimal medium. RNA was extracted twelve hours after transfer and cDNAs prepared."

ORIGIN

Query Match 3.2%; Score 61; DB 5; Length 487;  
Best Local Similarity 48.4%; Pred. No. 2.4e-05;  
Matches 240; Conservative 0; Mismatches 241; Indels 15; Gaps 2;

QY 614 GGATTGGGTGGTGGAGCTACTCTGGAGCTCATGGCTCTTCAGGATTGACTATAAGAG 673  
DB 1 GCATTGGATCAGCGGTGCTACTTTGATGTTTCGGTTCGACGGCTTCGACTCAACTT 60  
QY 674 CGACTGGCTCAGAACCAAGACCAATAGACTCTATGTGCGTTTTCGAGCTTACCGATGG 733  
DB 61 CGAGCGCATCCAACACACGAGGACAATGAGCTCTCAATGCGTATAAAGACATGTTTCAACT 120  
QY 734 GTTTCCTCTACCTTGGCTCGTTCAAGGCTATCATGTGGGATTTTGTACCTTACTTCCG 793  
DB 121 GGCTGTGTGCGAAGCGAGCTTCTGGCAGACGATGACAAACATTTACGCGCCGTTCTTGCT 180  
QY 794 AACTATGAACCGGAGACATGAGATAC---CTTTGACTCAAGGATTAGCAGTTTCCCGACG 850  
DB 181 ACGCTCTTCCCTACCCAGAAAGCAGCGACCGTGAACGCTGCCGAGAAATCATCCGCGC 240  
QY 851 AGTTGGGATCAGCTTATGGAGCAAAAGACAGCGCGTGTGTTGGCTCAGCTTCCGATCA 910  
DB 241 CGTCAAGCGGCGACATCATCCAGCACAAAGAGC-----GCAAGATCGAGGAGGG 288  
QY 911 GGCTGTTGATAAAAAAGGATGTTTCAAGTTCGGGATATCTTAAGTCTCTAGTGAGAGCAA 970

Db 289 CATGCGGACGGCAACCCCTACGAGCGCAAGGACATCTCTACCCCTCTCTCTCAAGTCGAA 348  
 Qy 971 CATCGCGCCCAACCTGCTGAATCTCAAAAGCTGTCCGATGAGGAGTACTCGCTCAGAT 1030  
 Db 349 CATGTCACCGACATCCGCGCCCGACGCGCATCCGAGCGCGACCTCTCTGACAACAT 408  
 Qy 1031 CAGTAACCTGTTATTTCTCTGATATGAATCTTCTGACAGTCTTCACATGATGTTTCA 1090  
 Db 409 CAACACGCTGCTCTGCTCCGCGGTCGACACCTGTCCTCGCGCTCACCTGAGCGCTCCA 468  
 Qy 1091 CCGACTCTCAGAAGAC 1106  
 Db 469 CTTCTCTGCGGAGCAC 484  
 RESULT 8  
 CD275427  
 LOCUS T143B04144F (FHIG:B) Axenic plate culture Paxillus involutus cDNA  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CD275427  
 VERSION CD275427.1 GI:34387473  
 KEYWORDS EST.  
 SOURCE Paxillus involutus  
 ORGANISM Paxillus involutus  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 Boletales; Paxillineae; Paxillaceae; Paxillus.  
 REFERENCE 1 (bases 1 to 400)  
 AUTHORS Johansson,T., Le Quere,A., Ahren,D., Soderstrom,B., Erlandsson,R.,  
 Lundeberg,J., Uhlen,M. and Tunlid,A.  
 TITLE Transcriptional responses of Paxillus involutus and Betula pendula  
 during formation of ectomycorrhizal root tissue  
 JOURNAL Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
 COMMENT Contact: Johansson,T.  
 Fungal-Host Interaction Group (FHIG)  
 Microbial Ecology, Institution of Ecology  
 Ecology Building, Lund University, SE-223 62 Lund, Sweden  
 Tel: +46 46 222 45 49  
 Fax: +46 46 222 41 58  
 Email: tomas.johansson@bioekol.lu.se  
 PCR Primers  
 FORWARD: P104 (5'-GGGAGCGCGCATGTGTT-3')  
 BACKWARD: P105 (5'-AGTGAGCTCGAATGCGGCC-3')  
 Seq primer: P104  
 High quality sequence stop: 400.  
 Location/Qualifiers  
 1..400  
 /organism="Paxillus involutus"  
 /mol\_type="mRNA"  
 /strain="ATCC 200175"  
 /db\_xref="taxon:71150"  
 /tissue\_type="Mycelium"  
 /dev\_stage="25 days of growth after transfer"  
 /lab\_host="Escherichia coli BM25.8"  
 /clone\_lib="(FHIG:B) Axenic plate culture"  
 /note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This  
 EST clone is originating from one of three cDNA libraries,  
 constructed for transcript profiling of the mycorrhizal  
 interaction between the basidiomycete Paxillus involutus  
 and Betula pendula (birch). One library represents the  
 developed and functional mycorrhizal root tissue  
 ('(FHIG:A) Ectomycorrhiza plate culture'), a second  
 library represents axenically grown fungus ('(FHIG:B)  
 Axenic plate culture') and a third library represents  
 axenically grown plants ('(FHIG:C) Axenic plate culture').  
 Libraries were analyzed in parallel and 3555 (FHIG:A),  
 3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20)  
 ESTs of >99bp have been deposited. The cDNA libraries were  
 constructed from total RNA using the SMART cDNA library  
 construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)  
 according to the manufacturer's instructions. Full-length  
 cDNAs were trimmed by SfiI, fractionated and directionally

ligated into (lambda)Triplex2 arms. The lambda library was  
 converted to a plasmid library via site-specific  
 recombination at loxP sites in a Cre+ strain (E. coli  
 BM25.8). Plasmid clones were randomly collected and  
 analysed by DNA sequencing using a plasmid-specific  
 forward primer (P104)."  
 ORIGIN  
 Query Match 3.0%; Score 58.8; DB 6; Length 400;  
 Best Local Similarity 59.6%; Pred. No. 9.4e-05;  
 Matches 99; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 Qy 1069 CAGTCTTGACATGGATGTTTACCGACTCTCAGAAGCAAAAGCGTTTCAGATAAATCTTC 1128  
 Db 52 CAGCGATAACTTGGCGCTGCATGAGTTGTGCTAGCCCTGGGATACAGAGGAAGCTCC 111  
 Qy 1129 GAGAAGAAATTTGCATGTCAGATCGACACGATATGCTACGCTAGACGAATTAATGCGTTGC 1188  
 Db 112 GGGAGGAATCTTCTTCTGTTCACACGGAAGCCCTCGATGGAGAACTCTCGGCGCTTC 171  
 Qy 1189 CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGTCTAGACCCCTCC 1234  
 Db 172 CTTACCTTGACACCGTGTGTGAAGGAGACTTTGCGTGTACACCCCTCC 217  
 RESULT 9  
 CD275197  
 LOCUS T143B01624F (FHIG:B) Axenic plate culture Paxillus involutus cDNA  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CD275197  
 VERSION CD275197.1 GI:34387243  
 KEYWORDS EST.  
 SOURCE Paxillus involutus  
 ORGANISM Paxillus involutus  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 Boletales; Paxillineae; Paxillaceae; Paxillus.  
 REFERENCE 1 (bases 1 to 317)  
 AUTHORS Johansson,T., Le Quere,A., Ahren,D., Soderstrom,B., Erlandsson,R.,  
 Lundeberg,J., Uhlen,M. and Tunlid,A.  
 TITLE Transcriptional responses of Paxillus involutus and Betula pendula  
 during formation of ectomycorrhizal root tissue  
 JOURNAL Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)  
 COMMENT Contact: Johansson,T.  
 Fungal-Host Interaction Group (FHIG)  
 Microbial Ecology, Institution of Ecology  
 Ecology Building, Lund University, SE-223 62 Lund, Sweden  
 Tel: +46 46 222 45 49  
 Fax: +46 46 222 41 58  
 Email: tomas.johansson@bioekol.lu.se  
 PCR Primers  
 FORWARD: P104 (5'-GGGAGCGCGCATGTGTT-3')  
 BACKWARD: P105 (5'-AGTGAGCTCGAATGCGGCC-3')  
 Seq primer: P104  
 High quality sequence stop: 317.  
 Location/Qualifiers  
 1..317  
 /organism="Paxillus involutus"  
 /mol\_type="mRNA"  
 /strain="ATCC 200175"  
 /db\_xref="taxon:71150"  
 /tissue\_type="Mycelium"  
 /dev\_stage="25 days of growth after transfer"  
 /lab\_host="Escherichia coli BM25.8"  
 /clone\_lib="(FHIG:B) Axenic plate culture"  
 /note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This  
 EST clone is originating from one of three cDNA libraries,  
 constructed for transcript profiling of the mycorrhizal  
 interaction between the basidiomycete Paxillus involutus  
 and Betula pendula (birch). One library represents the  
 developed and functional mycorrhizal root tissue  
 ('(FHIG:A) Ectomycorrhiza plate culture'), a second  
 library represents axenically grown fungus ('(FHIG:B)

Axenic plate culture') and a third library represents axenically grown plants ('(PHIG:C) Axenic plate culture'). Libraries were analyzed in parallel and 3555 (PHIG:A), 3964 (PHIG:B), and 2532 (PHIG:C) high-quality (PHRED 20) ESTs of >95bp have been deposited. The cDNA libraries were constructed from total RNA using the SMART cDNA library construction kit (#K1051-1, Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Full-length cDNAs were trimmed by SfiI, fractionated and directionally ligated into (lambda)Triplex2 arms. The lambda library was converted to a plasmid library via site-specific recombination at loxP sites in a Cre+ strain (S. coli BM25.8). Plasmid clones were randomly collected and analysed by DNA sequencing using a plasmid-specific forward primer (p104)."

## ORIGIN

Query Match 3.0%; Score 58; DB 6; Length 317;  
Best Local Similarity 55.4%; Pred. No. 0.00015;  
Matches 112; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1069 CAGTCTTGACATGGATGTTTCACCGACTCTCAGAGACAAGCGGTCAGGATAAATTC 1128  
Db 60 CAGGATAAATCTGGCGCTGCATGAGTTGCTGAGCCCTCGATGAGCAACTCTCGCGCTTC 119

QY 1129 GAGAAGAAATTTGTCAGATCGACACGGATATCCCTAGCTAGACGAATTAATGCGTTGC 1188  
Db 120 GCGAGGAATCTTCTTGTTGACACGGAAGCCCTCGATGAGCAACTCTCGCGCTTC 179

QY 1189 CTTATCTCGAAGCGTTTGTGAAGAGTCTCTCTGCTAGACCCCTCTAGTCCGTATGCTA 1248  
Db 180 CCTACCTTGACACCGTGGTGAAGAGACTTTGCGTGTGCACCCCTCCCTTTGGAGAGACCA 239

QY 1249 ACCCTGAATGCTTAAAGATGA 1270  
Db 240 TACGCTGCGGATGAAGAGCA 261

## RESULT 10

BU075377  
LOCUS BU075377 499 bp mRNA linear EST 29-SEP-2003  
DEFINITION laevis cDNA clone XL057f08 5', mRNA sequence.

## ACCESSION

BU075377 GI:17520293

## VERSION

EST.

## KEYWORDS

Xenopus laevis (African clawed frog)

## SOURCE

Xenopus laevis

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus; Xenopus.

1. (bases 1 to 499)

Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.

Expressed genes in X. laevis embryo

Unpublished (2001)

Contact: Tadasu Shin-i

Center for Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp

The information of this clone is available through the following URL.

http://xenopus.nibb.ac.jp.

Location/Qualifiers

1. .499

/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

/clone="XL057f08"

/tissue\_type="whole embryo"

## FEATURES

## source

1. .499  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL057f08"  
/tissue\_type="whole embryo"

## ORIGIN

Query Match 2.9%; Score 56.8; DB 4; Length 499;  
Best Local Similarity 53.8%; Pred. No. 0.00036;  
Matches 143; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

QY 996 CAAAGCTCTCCGATGAGGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCGTGATAT 1055  
Db 52 CAGAGGCTTCTGATGAGGATATACCGCTGAGGTGGACACATTATGTTTGGGACAT 111

QY 1056 GAAACTTCTTCGACAGCTCTTGACATGGATGTTTTCACCGACTCTCAGAGACAAGCCGTT 1115  
Db 112 GATACTACATCTAGTGGAAATATCTTGGATCCTACATTGTATGGCTACACATCCAGAGCA 171

QY 1116 CAGGATAAACTTCGAGAGAAATTTGTAGATCGACACGGATATGCTTACGCTAGAC 1172  
Db 172 CAGCAGAAATGCCAGGAGAGATCAATGAAGCTTTGGGGGAGCGCAACCAATGGAATGG 231

QY 1173 ---GAACCTTAATCGCTTGCTTATCTCGAAGCGTTTGTTAAGGAGTCTCTTCGTCTAGAC 1229  
Db 232 GATGACCTTAA TAGGATACCTTATACCAATGTCTATTAGGAAGTCTGCGCTTATAC 291

QY 1230 CCTCCTAGTCCGCTATGCTTAACCGTGA 1255  
Db 292 COTCCAGTCCCATCATGATGCCAGGA 317

## RESULT 11

BU0779027/c

LOCUS BU0779027 971 bp mRNA linear EST 10-DEC-2003

DEFINITION XGC-egg Xenopus tropicalis cDNA clone TEGG070n12 3', mRNA

sequence.

ACCESSION BU0779027

VERSION BU0779027.1

KEYWORDS GI:39686233

SOURCE EST.

ORGANISM Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus; Silurana.

1. (bases 1 to 971)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TEGG070n12.q1k17

Sequencing primer: T7

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from 5' of poly A+ RNA from egg.

5' end and NotI at the 3' end.

Vector: pCS107; Site1: EcoRI; Site2: NotI

Host: Escherichia coli XL1-blue.

Location/Qualifiers

1. .971

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TEGG070n12"

/dev\_stage="egg"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-egg"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from 5' of poly A+ RNA from egg

EcoRI-NotI cut cDNA was then ligated into pCS107 with

/dev\_stage="stage 25"

/clone\_lib="NIBB Mochii normalized Xenopus tailbud

library"



```

QY 995 TCAAAAGCTGTCGAGTGGAGGTACTCGCTCAGATCAGTAACCTGTTATTGCTGGATA 1054
Db 480 TAAAGCTTGACAGATGAGAAATTAATGGCTCAGTCCGTTATATTATATAGTCGCTGGATA 539
QY 1055 TGAATCTCTTCGACAGCTTTGACATGGATGTTTCAACGACTCTCAGAAAGACAAAGCGGT 1114
Db 540 TGAATCCACAGTACAGCAGCTGCTTACCTGTTTATTAATCTTGCCACCCCTGAGCT 599
QY 1115 TCAGGATAAATTCGAGAGAAATTT-----TGTCAGATCGACAGGATATGCTTACGCT 1168
Db 600 CCAGCAGAGACTGCGATGAGGAGATAGATTCATTTCTCCCTGACAGGCCAGCTCTACTTTA 659
QY 1169 AGACGAATTAATGCGTTGCTTATCTCGAAGCGTTTGTAAAGAGTCTCTTCTGCTAGA 1228
Db 650 TGACATTTCAATGCAGATGGAATATCTTGATGTTGTTATTCAGAGACCTTAGTGTTT 719
QY 1229 CCTCTCT 1235
Db 720 TCCCCCT 726

```

```

RESULT 14
AW600900
LOCUS DEFINITION
7C2 cDNA library of 4-day-old Eucalyptus globulus
bicostrata-Pisolithus tinctorius ectomycorrhiza Pisolithus
tinctorius cDNA 5' similar to hydrophobin, mRNA sequence.
ACCESSION
VERSION AW600900.1 GI:7304771
KEYWORDS EST.
SOURCE Pisolithus tinctorius
ORGANISM Pisolithus tinctorius
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Boletales; Sclerodermatineae; Pisolithaceae; Pisolithus.
REFERENCE
AUTHORS Voiblet,C., Duplessis,S., Encelet,N. and Martin,F.
TITLE 1 (bases 1 to 568)
Identification of symbiosis-regulated genes in Eucalyptus
Globulus-Pisolithus tinctorius ectomycorrhiza by differential
hybridization of arrayed cDNAs
JOURNAL Plant J. 25 (2), 181-191 (2001)
MEDLINE 21097016
PUBMED 11169194
COMMENT Contact: Martin FM
Equipe de Microbiologie Forestiere
Institut National de la Recherche Agronomique
Centre INRA de Nancy, 54280 Champenoux, France
Tel: +33 383 39 40 80
Fax: +33 383 39 40 69
Email: fmartin@nancy.inra.fr
Insert Length: 568 Std Error: 0.00
Seq primer: CapPinder 5'CDS
POLYA=Yes.

```

```

FEATURES
source
1. .568
Location/Qualifiers
/organism="Pisolithus tinctorius"
/mol_type="mRNA"
/strain="441"
/db_xref="taxon:37468"
/tissue_type="mantle and Hartig net hyphae"
/dev_stage="symbiotic, 4 days after contact"
/lab_host="Eucalyptus globulus bicostrata"
/clone_lib="cDNA library of 4-day-old Eucalyptus globulus
bicostrata-Pisolithus tinctorius ectomycorrhiza"
/note="Organ: ectomycorrhiza; Vector: pBluescript; Site 1:
EcoRI; The cDNA library was constructed from 1 ug total
ectomycorrhiza RNA using the SMART cDNA library
Construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)
according to the manufacturer's instructions. cDNAs were
cloned into EcoRI-digested pBluescript."

```

## ORIGIN

```

Query Match 2.7%; Score 51.2; DB 2; Length 568;
Best Local Similarity 58.8%; Pred. No. 0.013;

```

```

Matches 89; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 1762 AATGCCCTCTTTCTATCGATTGTTTCTCATACCCGCGAGGCGCTATGACTTCTACGTC 1821
Db 400 AATACATTTCAATTTTGGTGTGTTGCGGACAGATGAATTTGGATTGTTCTACTACGTA 459
QY 1822 GTCTATCGTCGCTCTCGAGCTCTTTTACCCCTATATATATTTCATCCGAAAAA 1881
Db 460 TTGACGGGTATAGAGGCTTAGCTTCTATTATAATAAATGCTTGGTTTCTCTAAAAA 519
QY 1882 AAAAAA 1913
Db 520 AAAAAA 551

```

```

RESULT 15
AW600865
LOCUS DEFINITION
6C8 cDNA library of 4-day-old Eucalyptus globulus
bicostrata-Pisolithus tinctorius ectomycorrhiza Pisolithus
tinctorius cDNA 5' similar to cross-pathway control wd-repeat
protein cpc-2, mRNA sequence.
ACCESSION
VERSION AW600865.1 GI:7304838
KEYWORDS EST.
SOURCE Pisolithus tinctorius
ORGANISM Pisolithus tinctorius
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Boletales; Sclerodermatineae; Pisolithaceae; Pisolithus.
REFERENCE
AUTHORS Voiblet,C., Duplessis,S., Encelet,N. and Martin,F.
TITLE 1 (bases 1 to 1296)
Identification of symbiosis-regulated genes in Eucalyptus
Globulus-Pisolithus tinctorius ectomycorrhiza by differential
hybridization of arrayed cDNAs
JOURNAL Plant J. 25 (2), 181-191 (2001)
MEDLINE 21097016
PUBMED 11169194
COMMENT Contact: Martin FM
Equipe de Microbiologie Forestiere
Institut National de la Recherche Agronomique
Centre INRA de Nancy, 54280 Champenoux, France
Tel: +33 383 39 40 80
Fax: +33 383 39 40 69
Email: fmartin@nancy.inra.fr
Insert Length: 1296 Std Error: 0.00
Seq primer: CapPinder 5'CDS
POLYA=Yes.

```

```

FEATURES
source
1. .1296
Location/Qualifiers
/organism="Pisolithus tinctorius"
/mol_type="mRNA"
/strain="441"
/db_xref="taxon:37468"
/tissue_type="mantle and Hartig net hyphae"
/dev_stage="symbiotic, 4 days after contact"
/lab_host="Eucalyptus globulus bicostrata"
/clone_lib="cDNA library of 4-day-old Eucalyptus globulus
bicostrata-Pisolithus tinctorius ectomycorrhiza"
/note="Organ: ectomycorrhiza; Vector: pBluescript; Site 1:
EcoRI; The cDNA library was constructed from 1 ug total
ectomycorrhiza RNA using the SMART cDNA library
Construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)
according to the manufacturer's instructions. cDNAs were
cloned into EcoRI-digested pBluescript."

```

## ORIGIN

```

Query Match 2.6%; Score 50.6; DB 2; Length 1296;
Best Local Similarity 76.5%; Pred. No. 0.025;
Matches 62; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 1833 CTCTGGACTCTCTTTACCTCATATATTATTCATCCGAAAAA 1892
Db 1150 CTCTGCCATATGCGTTGGGTTTCATAGCCATCCCCCGAAAAA 1209

```

QY 1893 AAAAAAAAAAGCGCGCTCGA 1913  
|||  
Db 1210 AAAAAAAAAAGCGCGCTCGA 1230  
|||

Search completed: December 14, 2004, 14:27:22  
Job time : 5824.92 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 01:52:58 ; Search time 941.936 Seconds  
(without alignments)  
11308.754 Million cell updates/sec

Title: US-10-066-007A-2

Perfect score: 1932  
Sequence: 1 gaattggcagcaggccacc.....agccggctcgtgcgaattc 1932

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/1/pubpna/US06\_PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

18: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*

20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1932	100.0	1932	14	US-10-066-007-2
2	353.2	18.3	3969	14	US-10-066-007-2
3	51.8	2.7	702	18	US-10-739-930-3162
4	49.8	2.6	333	18	US-10-425-115-54628
5	49.8	2.6	467	16	US-10-424-599-20949
6	49.4	2.6	1501	16	US-10-264-237-392
7	49.2	2.5	368	16	US-10-424-599-1793
8	48.4	2.5	657	16	US-10-424-599-25496
9	48.4	2.5	1458	18	US-10-690-991-1
10	48.4	2.5	1512	16	US-10-313-963A-55
11	48.4	2.5	2011	9	US-09-880-107-1586
12	48.4	2.5	2059	16	US-10-641-643-1062

13	48.4	2.5	2759	14	US-10-146-575-1	Sequence 1, Appli
14	48.4	2.5	2768	15	US-10-268-822-14	Sequence 14, Appl
15	48.4	2.5	2768	15	US-10-388-360-297	Sequence 297, App
16	48.4	2.5	2768	15	US-10-388-360-363	Sequence 363, App
17	48.4	2.5	2849	9	US-09-880-107-2110	Sequence 2110, Ap
18	48.2	2.5	312	9	US-09-960-352-4611	Sequence 4611, Ap
19	47.6	2.5	3059	9	US-09-925-297-268	Sequence 268, App
20	47.4	2.5	628	16	US-10-424-599-71313	Sequence 71313, A
21	47.4	2.5	1410	16	US-10-395-463-29	Sequence 29, Appl
22	47.4	2.5	1702	17	US-10-437-963-41374	Sequence 41374, A
23	47.4	2.5	2160	16	US-10-632-983-12	Sequence 12, Appl
24	47.2	2.4	1228	16	US-10-424-599-32438	Sequence 32438, A
25	47	2.4	218	9	US-09-933-797-162	Sequence 162, App
26	47	2.4	685	16	US-10-424-599-8656	Sequence 8656, Ap
27	46.8	2.4	709	16	US-10-424-599-75080	Sequence 75080, A
28	46.8	2.4	1081	18	US-10-425-115-155277	Sequence 155277, A
29	46.8	2.4	1151	16	US-10-424-599-23175	Sequence 106, App
30	46.8	2.4	1712	9	US-09-981-876-106	Sequence 106, App
31	46.8	2.4	1712	10	US-09-148-545-105	Sequence 105, App
32	46.8	2.4	1822	9	US-09-981-876-105	Sequence 105, App
33	46.8	2.4	1822	10	US-09-148-545-105	Sequence 33549, A
34	46.6	2.4	1091	16	US-10-424-599-33549	Sequence 72966, A
35	46.4	2.4	1419	16	US-10-424-599-72966	Sequence 106636, A
36	46.2	2.4	672	16	US-10-424-599-106636	Sequence 99744, A
37	46.2	2.4	710	16	US-10-424-599-99744	Sequence 15121, A
38	46.2	2.4	1194	16	US-10-424-599-15121	Sequence 29930, A
39	46	2.4	266	16	US-10-424-599-29930	Sequence 68827, A
40	46	2.4	414	15	US-10-424-599-64827	Sequence 3650, Ap
41	46	2.4	803	18	US-10-739-930-3650	Sequence 94220, A
42	46	2.4	940	16	US-10-424-599-94220	Sequence 111262, A
43	45.8	2.4	580	16	US-10-424-599-111262	Sequence 950, App
44	45.8	2.4	592	14	US-10-066-543-950	Sequence 54757, A
45	45.8	2.4	991	16	US-10-424-599-54757	

#### ALIGNMENTS

RESULT 1  
US-10-066-007-2  
Sequence 2, Application US/10066007  
Publication No. US20030077691A1  
GENERAL INFORMATION:  
APPLICANT: HOSHINO, Tatsuo  
APPLICANT: OJIMA, Kazuyuki  
TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
CURRENT APPLICATION NUMBER: US/10/066,007  
CURRENT FILING DATE: 2001-02-01  
PRIOR APPLICATION NUMBER: US/09/518,386  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: EP 99104668.1  
PRIOR FILING DATE: 1999-03-09  
PRIOR APPLICATION NUMBER: EP 00101666.6  
PRIOR FILING DATE: 2000-02-01  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1932  
TYPE: DNA  
ORGANISM: Phaffia rhodozyma  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (33)..(1706)  
NAME/KEY: polyA\_site  
LOCATION: (1871)  
NAME/KEY: mRNA  
LOCATION: (14)..(1891)  
US-10-066-007-2

Query Match 100.0%; Score 1932; DB 14; Length 1932;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	GAATTCGGCAGAGGCCACCTTCTCCATATGTTTCATCTTGGTCTGCTACAGGTG 60
Db 1	GAATTCGGCAGAGGCCACCTTCTCCATATGTTTCATCTTGGTCTGCTACAGGTG 60
QY 61	CTTTAGGCTGGCTGCTTCTCATGGGCATCCATAGCGTTCTTCAGTCTTTTACCTCGCTC 120
Db 61	CTTTAGGCTGGCTGCTTCTCATGGGCATCCATAGCGTTCTTCAGTCTTTTACCTCGCTC 120
QY 121	CGAGGCGATCTTCACTGATATACCTTCAGGCCCCGAATCATACCACTCTTTACAGGCA 180
Db 121	CGAGGCGATCTTCACTGATATACCTTCAGGCCCCGAATCATACCACTCTTTACAGGCA 180
QY 181	ATTTTATAGACATCTCTCAGCTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAAT 240
Db 181	ATTTTATAGACATCTCTCAGCTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAAAT 240
QY 241	ACGGAGCACCTCCGGTTGCTGGGATCGCTGGAGCACCGTCTTGAATCGACCGATC 300
Db 241	ACGGAGCACCTCCGGTTGCTGGGATCGCTGGAGCACCGTCTTGAATCGACCGATC 300
QY 301	CGAAAGTCTTCAACCATGTGATGAAGAGCCTACGACTATCCGAAACCTGGTATGCGCG 360
Db 301	CGAAAGTCTTCAACCATGTGATGAAGAGCCTACGACTATCCGAAACCTGGTATGCGCG 360
QY 361	CTCGAGTCTCAGAAATTCCTACCGAGATGGTGTGTCGGCGGAAGGTGAGCTCAT 420
Db 361	CTCGAGTCTCAGAAATTCCTACCGAGATGGTGTGTCGGCGGAAGGTGAGCTCAT 420
QY 421	AGCGACATCGAAGGATCATGATCCCTCTCTGTCCGCTCAGGCGCTTAAGTCGATGCTC 480
Db 421	AGCGACATCGAAGGATCATGATCCCTCTCTGTCCGCTCAGGCGCTTAAGTCGATGCTC 480
QY 481	CAATTTCTTAGAAAAGGATGGAATCTGTGCAAGAGATGATGGAGATCGCGCTGAGA 540
Db 481	CAATTTCTTAGAAAAGGATGGAATCTGTGCAAGAGATGATGGAGATCGCGCTGAGA 540
QY 541	AGGATATGCGCGTGGAGAGTCCGCGGTGAAAGAGCAAGCAACAGACTCGAGACCGAAG 600
Db 541	AGGATATGCGCGTGGAGAGTCCGCGGTGAAAGAGCAAGCAACAGACTCGAGACCGAAG 600
QY 601	GAGTCGATGAAGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 660
Db 601	GAGTCGATGAAGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 660
QY 661	TTGACTATAAGAGCGACTCGCTCCAGAACAGACCAATGAGCTCTATGCTGCTTTTGTG 720
Db 661	TTGACTATAAGAGCGACTCGCTCCAGAACAGACCAATGAGCTCTATGCTGCTTTTGTG 720
QY 721	GACTTACCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 780
Db 721	GACTTACCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 780
QY 781	TACCTTACTTCCGAATCTGAAACGGAGACATGAGATACCTTTGACTCAAGAGATTAGCAG 840
Db 781	TACCTTACTTCCGAATCTGAAACGGAGACATGAGATACCTTTGACTCAAGAGATTAGCAG 840
QY 841	TTTCCCGAGGTTGGATCGAGCTTATGGAGCAAGAGAGCGCGCTGCTGGCTCAG 900
Db 841	TTTCCCGAGGTTGGATCGAGCTTATGGAGCAAGAGAGCGCGCTGCTGGCTCAG 900
QY 901	CTTCCGATCAGGCTTGTGATAAAAGGATGTTCAAGGTCCGGATATCTTAAGTCTCCTAG 960
Db 901	CTTCCGATCAGGCTTGTGATAAAAGGATGTTCAAGGTCCGGATATCTTAAGTCTCCTAG 960
QY 961	TGAGAGCAACATCGCGCCAACTCGCTGATCTCAAAAGCTTCGATCAGAGGTAC 1020
Db 961	TGAGAGCAACATCGCGCCAACTCGCTGATCTCAAAAGCTTCGATCAGAGGTAC 1020
QY 1021	TCGCTCAGATCAGTAACCTGTTATTTGCTGGATATGAACCTTCTTCACAGTCTTGACAT 1080
Db 1021	TCGCTCAGATCAGTAACCTGTTATTTGCTGGATATGAACCTTCTTCACAGTCTTGACAT 1080

QY 1081	GGATCTTTACCGACTCTCAGAGCAAAAGCCGTTCCAGGATAAATTCGAGAGAAATTT 1140
Db 1081	GGATCTTTACCGACTCTCAGAGCAAAAGCCGTTCCAGGATAAATTCGAGAGAAATTT 1140
QY 1141	GTGAGATCGACACGATATGCTTACGCTAGACGAACTTAATGCGTTTCTTCTCGAAG 1200
Db 1141	GTGAGATCGACACGATATGCTTACGCTAGACGAACTTAATGCGTTTCTTCTCGAAG 1200
QY 1201	CGTTTGTATAGGAGTCTCTTCTGCTAGACCTCTCTAGTCGATGCTAAACCGTGAATGCT 1260
Db 1201	CGTTTGTATAGGAGTCTCTTCTGCTAGACCTCTCTAGTCGATGCTAAACCGTGAATGCT 1260
QY 1261	TAAAGGATGAAGACTTTCATCCCACTTSCCGAGCTCTGCTTGGTCGAGATGGTCCGTCA 1320
Db 1261	TAAAGGATGAAGACTTTCATCCCACTTSCCGAGCTCTGCTTGGTCGAGATGGTCCGTCA 1320
QY 1321	TCAACGAGTCCCGATCAAGAAAGGACGATGCTGCTTCCGTTGTTCAACATCAATC 1380
Db 1321	TCAACGAGTCCCGATCAAGAAAGGACGATGCTGCTTCCGTTGTTCAACATCAATC 1380
QY 1381	GTTCAAAAGTTCATTTATGGAAGATGCGAAGAAATTCAGACCGAGAGGTGCTTTGAGG 1440
Db 1381	GTTCAAAAGTTCATTTATGGAAGATGCGAAGAAATTCAGACCGAGAGGTGCTTTGAGG 1440
QY 1441	ACGTAAACAGACTCGCTCAACAGTATTGAAGACCCCTATGACACACAGGAGCTTTATCT 1500
Db 1441	ACGTAAACAGACTCGCTCAACAGTATTGAAGACCCCTATGACACACAGGAGCTTTATCT 1500
QY 1501	CTGACCCAGAGTCTGTTTGGTGGGATTTGCTCTGCGCGAGATGAAGGCTTCTTGT 1560
Db 1501	CTGACCCAGAGTCTGTTTGGTGGGATTTGCTCTGCGCGAGATGAAGGCTTCTTGT 1560
QY 1561	TTGTCACTCTCCGTCGGGTCAGTTCCGAGCCCATCATCTCATCCAGAGTACGAGCACA 1620
Db 1561	TTGTCACTCTCCGTCGGGTCAGTTCCGAGCCCATCATCTCATCCAGAGTACGAGCACA 1620
QY 1621	TCACCTTGATCATTTCCCGTCTCGAATCGTTGGTAGAGAGAGGGGTACAGATGC 1680
Db 1621	TCACCTTGATCATTTCCCGTCTCGAATCGTTGGTAGAGAGAGGGGTACAGATGC 1680
QY 1681	GTTTGCAGTCAAGCGCTCGAATGATGTTCTTCATATGTTAAGAGAGTCTTATAT 1740
Db 1681	GTTTGCAGTCAAGCGCTCGAATGATGTTCTTCATATGTTAAGAGAGTCTTATAT 1740
QY 1741	CTGAGATGTGATAGTAGCAATGCTTCTTGTATCGATTGTTCTCATACCGGGC 1800
Db 1741	CTGAGATGTGATAGTAGCAATGCTTCTTGTATCGATTGTTCTCATACCGGGC 1800
QY 1801	AGGCGCTACACTTACGTCGCTATCGTCGCTCGGACTCTCTTCTTACCTATAT 1860
Db 1801	AGGCGCTACACTTACGTCGCTATCGTCGCTCGGACTCTCTTCTTACCTATAT 1860
QY 1861	TATTCATCCGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCTCGAGCGGCT 1920
Db 1861	TATTCATCCGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCTCGAGCGGCT 1920
QY 1921	CGTCCCGAATTC 1932
Db 1921	CGTCCCGAATTC 1932

## RESULT 2

US-10-066-007-4  
; Sequence 4, Application US/10066007  
; Publication No. US20030077691A1  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/10/066,007

;; CURRENT FILING DATE: 2001-02-01  
;; PRIOR APPLICATION NUMBER: US/09/518,386  
;; PRIOR FILING DATE: 2000-03-03  
;; PRIOR APPLICATION NUMBER: EP 99104668.1  
;; PRIOR FILING DATE: 1999-03-09  
;; PRIOR APPLICATION NUMBER: EP 00101666.6  
;; PRIOR FILING DATE: 2000-02-01  
;; NUMBER OF SEQ ID NOS: 32  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 3969  
;; TYPE: DNA  
;; ORGANISM: Phaffia rhodozyma  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: (517)..(518)  
;; NAME/KEY: intron  
;; LOCATION: (784)..(898)  
;; NAME/KEY: intron  
;; LOCATION: (1016)..(1087)  
;; NAME/KEY: intron  
;; LOCATION: (1180)..(1302)  
;; NAME/KEY: intron  
;; LOCATION: (1518)..(1600)  
;; NAME/KEY: intron  
;; LOCATION: (1635)..(1723)  
;; NAME/KEY: intron  
;; LOCATION: (1867)..(1939)  
;; NAME/KEY: intron  
;; LOCATION: (2000)..(2081)  
;; NAME/KEY: intron  
;; LOCATION: (2182)..(2257)  
;; NAME/KEY: intron  
;; LOCATION: (2355)..(2431)  
;; NAME/KEY: intron  
;; LOCATION: (2543)..(2618)  
;; NAME/KEY: intron  
;; LOCATION: (2653)..(2742)  
;; NAME/KEY: intron  
;; LOCATION: (2815)..(2962)  
;; NAME/KEY: intron  
;; LOCATION: (3051)..(3113)  
;; NAME/KEY: intron  
;; LOCATION: (3172)..(3247)  
;; NAME/KEY: intron  
;; LOCATION: (3322)..(3398)  
;; NAME/KEY: intron  
;; LOCATION: (3424)..(3513)  
;; NAME/KEY: polyA\_site  
;; LOCATION: (3865)..(3866)  
;; NAME/KEY: intron  
;; LOCATION: (653)..(734)  
US-10-066-007-4

Query Match 18.3%; Score 353.2; DB 14; Length 3969;  
Best Local Similarity 99.2%; Pred. No. 4.1e-97;  
Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1514 TTGCTTTGGTTGGGATTTGCTGCGCGAGATGAGAGGCTTCTTTGTTGTCACCTCCG 1573  
DB 3508 TCGAGTGGTTGGGATTTGCTGCGCGAGATGAGAGGCTTCTTTGTTGTCACCTCCG 3567  
QY 1574 TCGGGTCCAGTTCGAGCCCATCATCTCTCATCCAGAGTACGAGCAGATCACCCTTGATCAT 1633  
DB 3568 TCGGGTCCAGTTCGAGCCCATCATCTCTCATCCAGAGTACGAGCAGATCACCCTTGATCAT 3627  
QY 1634 TTCCGCTCCGATGTTGGTGGTAGAGAGAGAGGGGTACGAGTCCGTTTCAGGTCAA 1693  
DB 3628 TTCCGCTCCGATGTTGGTGGTAGAGAGAGGGGTACGAGTCCGTTTCAGGTCAA 3687  
QY 1694 GCCGGTCCGATGATGATTTTCATATGTTAAGAGAAGTTCATATCTGAGAATGTGTG 1753  
DB 3688 GCCGGTCCGATGATGATTTTCATATGTTAAGAGAAGTTCATATCTGAGAATGTGTG 3747

QY 1754 ACTAGGACAATGCTCTTCTTGTATCGATTGTTTCTCATACCGGCGAGCGCTATGACT 1813  
DB 3748 ACTAGGACAATGCTCTTCTTGTATCGATTGTTTCTCATACCGGCGAGCGCTATGACT 3807  
QY 1814 TCTACGTCGCTCTATCGCTCTGACTCTCTTCTTACCTATATATTTCCATCCG 1871  
DB 3808 TCTACGTCGCTCTATCGCTCTGACTCTCTTCTTACCTATATATTTCCATCCG 3865

RESULT 3

US-10-739-930-3162  
; Sequence 3162, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 3162  
; LENGTH: 702  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-CLUSTER13983\_3  
US-10-739-930-3162

Query Match 2.7%; Score 51.8; DB 18; Length 702;  
Best Local Similarity 68.9%; Pred. No. 0.00011;  
Matches 71; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1812 CTCTACGTCGCTATCGCTCTGACTCTCTTCTTACCTATATATTTCCATCCG 1871  
DB 592 CATCTATGTTTGTGTAAGCTATGCCCGCATAGTATATATATATTTAGTCCAG 651  
QY 1872 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCTCGAG 1914  
DB 652 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCTCTAG 694

RESULT 4

US-10-425-115-54628/c  
; Sequence 54628, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 54628  
; LENGTH: 333  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_149820C.1  
US-10-425-115-54628

Query Match 2.6%; Score 49.8; DB 18; Length 333;  
Best Local Similarity 74.1%; Pred. No. 0.00029;  
Matches 63; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1824 CTATCGCTCTGACTCTTCTTACCTATATATTTCCATCCGAAAAAAAAA 1883  
DB 98 CTCGTGGTCTCGCTCTTCTTCTAGATAAATATTGCGCTATCCAAAAAAAAA 39

QY 1884 AAAAAAAAAAAAAAAAAAAGCGCGC 1908  
Db 38 AAAAAAAAAAAAAAAAAAAGCGCGC 14

## RESULT 5

US-10-424-599-20949  
; Sequence 1793; Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kowalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 20949  
; LENGTH: 467  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_118922C.1  
US-10-424-599-20949

Query Match 2.6%; Score 49.8; DB 16; Length 467;  
Best Local Similarity 96.2%; Pred. No. 0.00035;  
Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1861 TATTCATCCGAAAAAAAAAAAAAAAAAAGCGCGCTCGA 1913  
Db 401 TATTCATCCGAAAAAAAAAAAAAAAAAAGCGCGCTCGA 453

## RESULT 6

US-10-264-237-392  
; Sequence 392; Application US/10264237  
; Publication No. US20040009491A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P4131P1  
; CURRENT APPLICATION NUMBER: US/10/264,237  
; PRIOR FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/16450  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 60/205,515  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 2876  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 392  
; LENGTH: 1501  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (7)..(7)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-10-264-237-392

Query Match 2.6%; Score 49.4; DB 16; Length 1501;  
Best Local Similarity 53.2%; Pred. No. 0.00092;  
Matches 74; Conservative 1; Mismatches 42; Indels 0; Gaps 0;

QY 1785 TTTCATACCGCGCGCGCTATGACTTCTACGTCGTCTATCGTCGCTGACTCTC 1844  
Db 1370 TTTTCCTTGCTGACTTCTTGTGCACAGACTGCATAGTTGTGACGCTTGACTATC 1429

QY 1845 TTCTTACCCCTATATATTATTCATCCGAAAAAAAAAAAAAAAAAAAAAAAAA 1901

Db 1430 TTTTGAATAAGATTGATTTTAAACAAAAAAAAAAAAAAAAAAAAAAAAA 1486

## RESULT 7

US-10-424-599-1793  
; Sequence 1793; Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kowalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 1793  
; LENGTH: 368  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_101618C.1  
US-10-424-599-1793

Query Match 2.5%; Score 49.2; DB 16; Length 368;  
Best Local Similarity 87.1%; Pred. No. 0.00047;  
Matches 54; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1953 CTATATATTATTCATCCGAAAAAAAAAAAAAAAAAAGCGCGCTCG 1912  
Db 299 CAATAGATACCTTATCCGAAAAAAAAAAAAAAAAAAGCGCGCTCT 358

QY 1913 AG 1914  
Db 359 AG 360

## RESULT 8

US-10-424-599-25496/c  
; Sequence 25496; Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kowalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 25496  
; LENGTH: 657  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123024C.1  
US-10-424-599-25496

Query Match 2.5%; Score 48.4; DB 16; Length 657;  
Best Local Similarity 74.4%; Pred. No. 0.0012;  
Matches 61; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1833 CTCGGACTCTCTTACCCCTATATATTATTCATCCGAAAAAAAAAAAAAAAAA 1892  
Db 91 CTCGAGATCTCTGAAACAGATTCAAGATGAAGAGCAAAAAAAAAAAAAAAAAA 32

QY 1593 AAAAAAAAAAGCGCGCTCGAG 1914

```
Db      31 AAAAAAAAAAGCGCGCTCTAG 10

RESULT 9
US-10-690-991-1
; Sequence 1, Application US/10690991
; Publication No. US20040243319A1
; GENERAL INFORMATION:
; APPLICANT: Tickle, Ian J
; APPLICANT: Vonthein, Clemens
; APPLICANT: Williams, Pamela A
; APPLICANT: Jhoti, Harren
; APPLICANT: Kirtan, Stewart Brian
; TITLE OF INVENTION: Crystal structure of cytochrome P450
; FILE REFERENCE: 620-282
; CURRENT APPLICATION NUMBER: US/10/690,991
; CURRENT FILING DATE: 2004-04-13
; PRIOR APPLICATION NUMBER: US 60/421,063
; PRIOR FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: PCT/GB02/02668
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 10/221,036
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: US 60/479,448
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Codes for SEQ ID NO: 2, a human 3A4 P450 protein truncated in its
; OTHER INFORMATION: N-terminal region to delete the hydrophobic trans-membrane
; OTHER INFORMATION: domain, and the region replaced by a short N-terminal sequence.
US-10-690-991-1

Query Match      2.5%; Score 48.4; DB 18; Length 1458;
Best Local Similarity 50.7%; Pred. No. 0.0018;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
Db 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
QY 797 AAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 856
Db 797 AAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 856
QY 1057 AAACCTTCTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116
Db 1057 AAACCTTCTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116
QY 857 AAACCCAGAGCGAGTCTCTCCTCTCATTATGATGAACCTGGCCACTCACCCCTGATGTC 916
Db 857 AAACCCAGAGCGAGTCTCTCCTCTCATTATGATGAACCTGGCCACTCACCCCTGATGTC 916
QY 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTTACGCTAGACGAAC 1176
Db 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTTACGCTAGACGAAC 1176
QY 917 AGCAGAACTGCAGAGGAATGATGCAGTGTACCAATAGGACACCCACCTATG 976
Db 917 AGCAGAACTGCAGAGGAATGATGCAGTGTACCAATAGGACACCCACCTATG 976
QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTCGCTAGACC 1230
Db 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTCGCTAGACC 1230
QY 977 ATACTGTGTACAGATGAGTATCTTGACATGCTGTTTACCAATAGGACACCCACCTATG 1036
Db 977 ATACTGTGTACAGATGAGTATCTTGACATGCTGTTTACCAATAGGACACCCACCTATG 1036
QY 1231 CTCTAGTCGATGCTTACCGTGAATGCTTAAAGATGAGACTTCA 1278
Db 1231 CTCTAGTCGATGCTTACCGTGAATGCTTAAAGATGAGACTTCA 1278
QY 1037 CAATTGCTATGACTTGGAGGGTCTGCAAAAAGATGTTGAGATCA 1084
Db 1037 CAATTGCTATGACTTGGAGGGTCTGCAAAAAGATGTTGAGATCA 1084

RESULT 10
US-10-313-963A-55
; Sequence 55, Application US/10313963A
; Publication No. US2004002078A1
; GENERAL INFORMATION:
; APPLICANT: Boutell, Jonathan
; APPLICANT: Godber, Benjamin
; APPLICANT: Hart, Darren
; APPLICANT: Blackburn, Jonathan
; TITLE OF INVENTION: Arrays
; FILE REFERENCE: KIL-001
US-10-313-963A-55

Query Match      2.5%; Score 48.4; DB 9; Length 2011;
Best Local Similarity 50.7%; Pred. No. 0.0022;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
Db 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
QY 926 AAGCTCTGTCGATGAGAGGTCTGCAAAAAGATGTTGAGATCA 985
Db 926 AAGCTCTGTCGATGAGAGGTCTGCAAAAAGATGTTGAGATCA 985
```

```
; CURRENT APPLICATION NUMBER: US/10/313,963A
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US 60/335,806
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: US 60/410,815
; PRIOR FILING DATE: 2002-09-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-313-963A-55

Query Match      2.5%; Score 48.4; DB 16; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.0019;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
Db 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
QY 863 AAGCTCTGTCGATGAGAGGTCTGGAGCTCGGCCCAATCAATATCTTTATTTTGGCTATG 922
Db 863 AAGCTCTGTCGATGAGAGGTCTGGAGCTCGGCCCAATCAATATCTTTATTTTGGCTATG 922
QY 1057 AAACCTTCTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116
Db 1057 AAACCTTCTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116
QY 923 AAACCCAGAGCGAGTCTCTCCTCTCATTATGATGAACCTGGCCACTCACCCCTGATGTC 982
Db 923 AAACCCAGAGCGAGTCTCTCCTCTCATTATGATGAACCTGGCCACTCACCCCTGATGTC 982
QY 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTTACGCTAGACGAAC 1176
Db 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTTACGCTAGACGAAC 1176
QY 983 AGCAGAACTGCAGAGGAATGATGAGTGTATACCAATAGGACACCCACCTATG 1042
Db 983 AGCAGAACTGCAGAGGAATGATGAGTGTATACCAATAGGACACCCACCTATG 1042
QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTCGCTAGACC 1230
Db 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTCGCTAGACC 1230
QY 1043 ATACTGTGTACAGATGAGTATCTTGACATGCTGTTTACCAATAGGACACCCACCTATG 1102
Db 1043 ATACTGTGTACAGATGAGTATCTTGACATGCTGTTTACCAATAGGACACCCACCTATG 1102
QY 1231 CTCTAGTCGATGCTTACCGTGAATGCTTAAAGATGAGACTTCA 1278
Db 1231 CTCTAGTCGATGCTTACCGTGAATGCTTAAAGATGAGACTTCA 1278
QY 1103 CAATTGCTATGACTTGGAGGGTCTGCAAAAAGATGTTGAGATCA 1150
Db 1103 CAATTGCTATGACTTGGAGGGTCTGCAAAAAGATGTTGAGATCA 1150

RESULT 11
US-09-880-107-1586
; Sequence 1586, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1586
; LENGTH: 2011
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D00003
US-09-880-107-1586

Query Match      2.5%; Score 48.4; DB 9; Length 2011;
Best Local Similarity 50.7%; Pred. No. 0.0022;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
Db 997 AAAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056
QY 926 AAGCTCTGTCGATGAGAGGTCTGCAAAAAGATGTTGAGATCA 985
Db 926 AAGCTCTGTCGATGAGAGGTCTGCAAAAAGATGTTGAGATCA 985
```

QY 1057 AAACCTTCTTCGACAGCTTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116  
| | | | |  
Db 986 AAACCCAGGACAGTGTCTCTCTTCATATGATGAACCTGGCCACTCACCCCTGATGTC 1045  
| | | | |  
QY 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTACGCTAGACGAAC 1176  
| | | | |  
Db 1046 AGCAGAACTGCAGAGGAAATTTGATGCGAGTTTACCCCAATAGGCACCCACCCTATG 1105  
| | | | |  
QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGCTAGACC 1230  
| | | | |  
Db 1106 ATACTGTGCTACAGATGGAGTATCTTGACATGCTGTGTGATGAACGCTCAGATTATTC 1165  
| | | | |  
QY 1231 CTCCTAGTCGCTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
| | | | |  
Db 1166 CAATTGCTATGAGACTTGAAGGGTCTGCACAAAAGATGTTGAGATCA 1213  
| | | | |

## RESULT 12

US-10-641-643-1062  
; Sequence 1062, Application US/10641643  
; Publication No. US20040077003A1  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; Susan G. Stuart  
; Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL  
; GENE EXPRESSION

NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/641.643  
FILING DATE: 14-Aug-2003  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 1062:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2059 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: g181373  
SEQUENCE DESCRIPTION: SEQ ID NO: 1062 :  
US-10-641-643-1062

Query Match 2.5%; Score 48.4; DB 16; Length 2059;  
Best Local Similarity 50.7%; Pred. No. 0.0022;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;  
QY 997 AAAAGCTGTCGATGAGAGGCTGCTCAGATCAGTAACTGTTATTCCTGATG 1056  
| | | | |

Db 953 AAGCTCTGTCGATCGAGCTCGTGGCCCAATCAATATATCTTTTGTGGCTATG 1012  
| | | | |  
QY 1057 AAACCTTCTTCGACAGCTTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116  
| | | | |  
Db 1013 AAACCCAGGACAGTGTCTCTCTTCATATGATGAACCTGGCCACTCACCCCTGATGTC 1072  
| | | | |  
QY 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTACGCTAGACGAAC 1176  
| | | | |  
Db 1073 AGCAGAACTGCAGAGGAAATTTGATGCGAGTTTACCCCAATAGGCACCCACCCTATG 1132  
| | | | |  
QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGCTAGACC 1230  
| | | | |  
Db 1133 ATACTGTGCTACAGATGGAGTATCTTGACATGCTGTGTGATGAACGCTCAGATTATTC 1192  
| | | | |  
QY 1231 CTCCTAGTCGCTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
| | | | |  
Db 1193 CAATTGCTATGAGACTTGAAGGGTCTGCACAAAAGATGTTGAGATCA 1240  
| | | | |

## RESULT 13

US-10-146-575-1  
; Sequence 1, Application US/10146575  
; Publication No. US20030059800A1  
; GENERAL INFORMATION:  
; APPLICANT: Lichter, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/10/146.575  
; CURRENT FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: US/09/144.367  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2759  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (70)...(1581)  
; OTHER INFORMATION: Human CYP3A4 cDNA reference sequence  
US-10-146-575-1

Query Match 2.5%; Score 48.4; DB 14; Length 2759;  
Best Local Similarity 50.7%; Pred. No. 0.0026;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGAGGCTGCTCAGATCAGTAACTGTTATTCCTGATG 1056  
| | | | |  
Db 932 AAGCTCTGTCGATGTCGAGCTCGTGGCCCAATCAATATCTTTATTTTGTGGCTATG 991  
| | | | |

QY 1057 AAACTTCTTCGACAGCTTTGACATGGATGTTTCAACCGACTCTCAGAGACAAAGCCGTTTC 1116  
| | | | |  
Db 992 AAACCCAGGACAGTGTCTCTCTTCATATGATGAACCTGGCCACTCACCCCTGATGTC 1051  
| | | | |

QY 1117 AGGATAAACTTCGAGAGAAATTTGTCAGATCGACACGGATATGCTACGCTAGACGAAC 1176  
| | | | |  
Db 1052 AGCAGAACTGCAGAGGAAATTTGATGAGTTTTTACCANATAGGCACCCACCCTATG 1111  
| | | | |

QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGCTAGACC 1230  
| | | | |  
Db 1112 ATACTGTGCTACAGATGGAGTATCTTGACATGCTGTGATGAATGAACGCTCAGATTATTC 1171  
| | | | |

QY 1231 CTCCTAGTCGCTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
| | | | |  
Db 1172 CAATTGCTATGAGACTTGAAGGGTCTGCACAAAAGATGTTGAGATCA 1219  
| | | | |

## RESULT 14

US-10-268-822-14  
; Sequence 14, Application US/10268822  
; Publication No. US20030150004A1

Query Match	2.5%;	Score	48.4;	DB	15;	Length	2768;
Best Local Similarity	50.7%;	Pred. No.	0.0026;				
Matches	146;	Conservative	0;	Mismatches	136;	Indels	6;
Gaps							

  

Qy	997	AAAGCTGTCCGATGAGAGGTA	CTCGCTCAGATCAGTAACCTGTATTTCGTGGATG	1056
Db	967	AAGCTGTCCGATCTGGAGCT	GTGGCCAAATCAATTATCTTTATTTTGCTGCTATG	1026
Qy	1057	AAACTTCTTCGACAGCTCTTCACATGGATG	TGTTTACCAGACTCTCAGAAGACAAGCCGTT	1116
Db	1027	AAACCGAGCAGTGTTCTCTCTTCATATG	TATGAACCTGGCCACTCAGCCCTGATGCC	1086
Qy	1117	AGGATAAACTTCGAGAAGAAATTTGT	CAGATCGACACGGATATGCTTACCTAGACGAAC	1176
Db	1087	AGCAGAACTGCAGGAGGAAATTTGATG	CACTTTTACCCAAATAGGCACACCCACTTATG	1146
Qy	1177	TTAATGCGTTGC-----CTTATCTCGA	ACGTTTGTTAAGAGTCTCTTCGCTAGACC	1230
Db	1147	ATACTGTGCTACAGATGGAGTATCTT	GACATGTGGTGAATGAACCGCTCAGATTATTC	1206
Qy	1231	CTCCTAGTCGGTATGCTAACCGTGAAT	CTCTTAAGAGATGAAGACTTCA	1278
Db	1207	CAATTGCTATGACACTTGAGAGGGT	CTGCAAAAAGATGTTTGAGATCA	1254

```

: GENERAL INFORMATION:
: APPLICANT: Moore, David
: APPLICANT: Wei, Ping
: APPLICANT: Chua, Steven
: TITLE OF INVENTION: Screening Systems and Methods for Identifying Modulators of xenob
: TITLE OF INVENTION: Metabolism
: FILE REFERENCE: P02729US2
: CURRENT APPLICATION NUMBER: US/10/268,822
: CURRENT FILING DATE: 2002-10-10
: PRIOR APPLICATION NUMBER: PCT/US 01/29672
: PRIOR FILING DATE: 2001-09-21
: PRIOR APPLICATION NUMBER: US 10/219,590
: PRIOR FILING DATE: 2002-08-15
: NUMBER OF SEQ ID NOS: 27
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 2768
: TYPE: DNA
: ORGANISM: Human
: US-10-268-822-14

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 01:23:48 ; Search time 158.79 Seconds  
(without alignments)  
8648.176 Million cell updates/sec

Title: US-10-066-007A-2

Perfect score: 1932

Sequence: 1 gaattcgccagcagccacc.....agccggctcgtgcgaattc 1932

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1932	100.0	1932	3	US-09-518-386B-2
2	353.2	18.3	3969	3	US-09-518-386B-4
3	48.4	2.5	1512	4	US-08-277-031B-4
4	48.4	2.5	2059	4	US-09-023-655-1062
5	48.4	2.5	2759	4	US-09-144-367-1
6	47.4	2.5	2160	4	US-09-716-129-12
7	46.8	2.4	1712	4	US-09-148-545-106
8	46.8	2.4	1822	4	US-09-148-545-105
9	45	2.3	1831	3	US-08-948-564-15
10	44.8	2.3	1801	4	US-09-709-103-3
11	44.8	2.3	1801	4	US-09-439-410A-3
12	44	2.3	3828	4	US-09-221-013A-7
13	43.6	2.3	352	4	US-09-144-367-8
14	42.8	2.2	1333	3	US-09-372-422A-9
15	42.8	2.2	2407	3	US-09-370-807-7
16	42.8	2.2	2407	3	US-09-921-259-7
17	42.6	2.2	246240	2	US-08-724-394A-20
18	42.6	2.2	246240	2	US-08-724-394A-21
19	42.6	2.2	246240	2	US-08-724-394A-22
20	42.4	2.2	396	4	US-09-640-173-19
21	42.4	2.2	396	4	US-09-713-550-19
22	42.4	2.2	396	4	US-09-825-294-19
23	42.4	2.2	396	4	US-09-970-966-19
24	42.2	2.2	47	2	US-08-778-494B-114
25	42.2	2.2	323	4	US-09-621-976-10374
26	41.6	2.2	2202	4	US-09-396-149-3
27	41.6	2.2	2691	4	US-10-101-464A-837

28	41.4	2.1	1126	4	US-09-389-956-7	Sequence 7, Appli
29	41.4	2.1	1507	4	US-09-453-323-1	Sequence 1, Appli
C 30	41.4	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
C 31	41.2	2.1	1024	4	US-09-328-475C-107	Sequence 107, App
C 32	40.8	2.1	44848	4	US-09-435-733-42	Sequence 42, Appl
C 33	40.8	2.1	44848	4	US-09-988-113-42	Sequence 13, Appl
34	40.6	2.1	1883	4	US-09-419-679-13	Sequence 8, Appli
35	40.2	2.1	947	4	US-09-673-395A-8	Sequence 48, Appl
C 36	40.2	2.1	5152	4	US-10-204-708-48	Sequence 11009, A
37	40	2.1	155	4	US-09-621-976-11009	Sequence 1, Appli
38	40	2.1	1260	4	US-09-674-741-5	Sequence 1405, Ap
39	40	2.1	2581	2	US-09-013-634-1	Sequence 56, Appl
40	39.8	2.1	1599	4	US-09-023-655-1405	Sequence 16, Appl
41	39.8	2.1	2099	4	US-09-800-729-56	Sequence 16, Appl
42	39.6	2.0	214	4	US-09-621-976-9843	Sequence 16, Appl
C 43	39.6	2.0	396	4	US-09-640-173-16	Sequence 16, Appl
C 44	39.6	2.0	396	4	US-09-713-550-16	Sequence 16, Appl
C 45	39.6	2.0	396	4	US-09-825-294-15	Sequence 16, Appl

## ALIGNMENTS

### RESULT 1

US-09-518-386B-2

; Sequence 2, Application US/09518386B

; Patent No. 6365386

; GENERAL INFORMATION:

; APPLICANT: HOSHINO, Tatsuo

; APPLICANT: OJIMA, Kazuyuki

; APPLICANT: SETOUCHI, Yutaka

; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE

; FILE REFERENCE: ASTAXANTHIN SYNTHETASE

; CURRENT APPLICATION NUMBER: US/09/518,386B

; CURRENT FILING DATE: 2000-03-03

; PRIOR APPLICATION NUMBER: EP 99104668.1

; PRIOR FILING DATE: 1999-03-09

; PRIOR APPLICATION NUMBER: EP 00101666.6

; PRIOR FILING DATE: 2000-02-01

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; TYPE: DNA

; LENGTH: 1932

; ORGANISM: Phaffia rhodozyma

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (33)..(1706)

; NAME/KEY: polyA site

; LOCATION: (1871)

; NAME/KEY: mRNA

; LOCATION: (14)..(1891)

US-09-518-386B-2

Query Match 100.0%; Score 1932; DB 3; Length 1932;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCGCGCAGGCGCACCTACTTCTCCATATGTTTCATCTTGGTCTTGTCTCAGAGTG 60

Db 1 GAATTCGCGCAGGCGCACCTACTTCTCCATATGTTTCATCTTGGTCTTGTCTCAGAGTG 60

Qy 61 CTTTAGGCGCTGGCTGCTTTCTCAAGGATCCATAGCGCTTCTTACGCTTTACCTCGCTC 120

Db 61 CTTTAGGCGCTGGCTGCTTTCTCAAGGATCCATAGCGCTTCTTACGCTTTACCTCGCTC 120

Qy 121 CGAGCGCATCTTCACTGTATTAACCTTCAGGCGCCGATCATACCACTTACTTTACAGGCA 180

Db 121 CGAGCGCATCTTCACTGTATTAACCTTCAGGCGCCGATCATACCACTTACTTTACAGGCA 180

Qy 181 ATTTTITAGACATCTCTCAGCTCGTGAAGAGCATCCGAGTACAGAGAAAAT 240

Db 181 ATTTTITAGACATCTCTCAGCTCGTGAAGAGCATCCGAGTACAGAGAAAAT 240

Qy 181 ATTTTITAGACATCTCTCAGCTCGTGAAGAGCATCCGAGTACAGAGAAAAT 240

Db 181 ATTTTITAGACATCTCTCAGCTCGTGAAGAGCATCCGAGTACAGAGAAAAT 240

applied

241 ACAGAGACCCCTCCGCTTGGTGGATCGCTGGAGCACCCGCTCTTGAACCTGACCGATC 300  
Db |||||  
241 ACAGAGACCCCTCCGCTTGGTGGATCGCTGGAGCACCCGCTCTTGAACCTGACCGATC 300  
Qy |||||  
301 CGAAGCTCTCAACCATGTGATGAAGAAGCTTACGACTATCCGAACCTGGTATGGCCG 360  
Db |||||  
301 CGAAGCTCTCAACCATGTGATGAAGAAGCTTACGACTATCCGAACCTGGTATGGCCG 360  
Qy |||||  
361 CTCGAGTCTCAGAAATGCTACCCGAGATGGTGTGTTACGGCGGAAGGTGAAGCTCAT 420  
Db |||||  
361 CTCGAGTCTCAGAAATGCTACCCGAGATGGTGTGTTACGGCGGAAGGTGAAGCTCAT 420  
Qy |||||  
421 AGCGACATCGAAGATCATGATCCCTCTCTGCTCGCTCAGGCGGTTAAGTCCATGGTCC 480  
Db |||||  
421 AGCGACATCGAAGATCATGATCCCTCTCTGCTCGCTCAGGCGGTTAAGTCCATGGTCC 480  
Qy |||||  
481 CAATTTTCTTAGAAAAAGGTAAGAACTTGTGCAAGATGATGGAGGATGCGGCTGAGA 540  
Db |||||  
481 CAATTTTCTTAGAAAAAGGTAAGAACTTGTGCAAGATGATGGAGGATGCGGCTGAGA 540  
Qy |||||  
541 AGGATATGCGCGTGGAGAGTCCGCGGTGAAGAGGCAACAGACTCGAGACCGAAG 600  
Db |||||  
541 AGGATATGCGCGTGGAGAGTCCGCGGTGAAGAGGCAACAGACTCGAGACCGAAG 600  
Qy |||||  
601 GAGTCGATGTAAGGATTTGGTTCGAGTACTCTGAGCTCATGGCTCTTGCAGGAT 660  
Db |||||  
601 GAGTCGATGTAAGGATTTGGTTCGAGTACTCTGAGCTCATGGCTCTTGCAGGAT 660  
Qy |||||  
661 TTGACTATAAGAGCGACTCGCTCAGAACCAAGACCAATGAGCTCTATGTCGCTTTGTCG 720  
Db |||||  
661 TTGACTATAAGAGCGACTCGCTCAGAACCAAGACCAATGAGCTCTATGTCGCTTTGTCG 720  
Qy |||||  
721 GACTTACCGATGGTTCCTCCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTG 780  
Db |||||  
721 GACTTACCGATGGTTCCTCCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTG 780  
Qy |||||  
781 TACTTACTTCCGAATATGAACCGAGACATGATACCTTTGACTCAAGGATTAGCAG 840  
Db |||||  
781 TACTTACTTCCGAATATGAACCGAGACATGATACCTTTGACTCAAGGATTAGCAG 840  
Qy |||||  
841 TTTCGCGAGGATGGGATCGAGCTTATGAGCAAAAGAGACGCGCTGCTGGCTCAG 900  
Db |||||  
841 TTTCGCGAGGATGGGATCGAGCTTATGAGCAAAAGAGACGCGCTGCTGGCTCAG 900  
Qy |||||  
901 CTTCCGATCAGGCTGTTGATAAAAAGGATGTTCAAGGTCGGGATATCTAAGTCTCCTAG 960  
Db |||||  
901 CTTCCGATCAGGCTGTTGATAAAAAGGATGTTCAAGGTCGGGATATCTAAGTCTCCTAG 960  
Qy |||||  
961 TGAGAGCAAACTCCGCCCAACCTGCTGAATCTCAAAAGCTGCTCGATCAGGAGTAC 1020  
Db |||||  
961 TGAGAGCAAACTCCGCCCAACCTGCTGAATCTCAAAAGCTGCTCGATCAGGAGTAC 1020  
Qy |||||  
1021 TCGCTCAGATCAGTAACCTGTTATTTGCTGGATATGAACCTTTCTCGACAGCTTTGACAT 1080  
Db |||||  
1021 TCGCTCAGATCAGTAACCTGTTATTTGCTGGATATGAACCTTTCTCGACAGCTTTGACAT 1080  
Qy |||||  
1081 GGATGTTTCAACGATCTCAGAAACAAAGCGGTTGAGGATAAATCTTCGAGAAAGATTT 1140  
Db |||||  
1081 GGATGTTTCAACGATCTCAGAAACAAAGCGGTTGAGGATAAATCTTCGAGAAAGATTT 1140  
Qy |||||  
1141 GTCAGATCGACACGATATGCTAGCTAGACGAACTTAATGCGTTCGCTTATCTCGAAG 1200  
Db |||||  
1141 GTCAGATCGACACGATATGCTAGCTAGACGAACTTAATGCGTTCGCTTATCTCGAAG 1200  
Qy |||||  
1201 CGTTTGTAAAGGATCTCTTCTGCTAGACCTCTCTAGTCCGTATGCTAACCGTGAATGCT 1260  
Db |||||  
1201 CGTTTGTAAAGGATCTCTTCTGCTAGACCTCTCTAGTCCGTATGCTAACCGTGAATGCT 1260  
Qy |||||  
1261 TAAAGGATGAAGACTTCAATCCACCTGCGAGCGCTGCTATTTGGTCGAGATGGTTCGGTCA 1320  
Db |||||  
1261 TAAAGGATGAAGACTTCAATCCACCTGCGAGCGCTGCTATTTGGTCGAGATGGTTCGGTCA 1320

1321 TCAACGAGTCCGATCACGAAAGAAACGATGCTCATGCTTCGTTTCAACATCAATC 1380  
Db |||||  
1321 TCAACGAGTCCGATCACGAAAGAAACGATGCTCATGCTTCGTTTCAACATCAATC 1380  
Qy |||||  
1381 GTTCAAAAGTTCATTTATGGAGAAATGCAAGAAATTCAGACCCGAGAGGTGGCTTCAAG 1440  
Db |||||  
1381 GTTCAAAAGTTCATTTATGGAGAAATGCAAGAAATTCAGACCCGAGAGGTGGCTTCAAG 1440  
Qy |||||  
1441 ACGTAAACAGATCGCTCAACAGTATTGAAGCACCTTATGACACCCAGCGAGCTTTATCT 1500  
Db |||||  
1441 ACGTAAACAGATCGCTCAACAGTATTGAAGCACCTTATGACACCCAGCGAGCTTTATCT 1500  
Qy |||||  
1501 CTGGACCCAGAGCTTCGTTGGTGGCGATTTGCTGTCGCCGAGATGAAGGCTTCTTGT 1560  
Db |||||  
1501 CTGGACCCAGAGCTTCGTTGGTGGCGATTTGCTGTCGCCGAGATGAAGGCTTCTTGT 1560  
Qy |||||  
1561 TTGTCACCTCCGTCGGGTCAGTTCCAGCCCATCATCTCTCATCCAGAGTACGAGCACA 1620  
Db |||||  
1561 TTGTCACCTCCGTCGGGTCAGTTCCAGCCCATCATCTCTCATCCAGAGTACGAGCACA 1620  
Qy |||||  
1621 TCACCTTGCATCATTTCCGTCCTCGAATCGTTGGTAGAGAAAGAGGGTACGAGATGC 1680  
Db |||||  
1621 TCACCTTGCATCATTTCCGTCCTCGAATCGTTGGTAGAGAAAGAGGGTACGAGATGC 1680  
Qy |||||  
1681 GTTTCAGGTCAGCCGTCGAATGAGTTGATTTCTTCAATGTTAAAGAGAGTTCTATAT 1740  
Db |||||  
1681 GTTTCAGGTCAGCCGTCGAATGAGTTGATTTCTTCAATGTTAAAGAGAGTTCTATAT 1740  
Qy |||||  
1741 CTGAGAAATGTGACTAGGACAATGCCCTTCTTTGTCATCGATTTGTTCTCATACCCGGC 1800  
Db |||||  
1741 CTGAGAAATGTGACTAGGACAATGCCCTTCTTTGTCATCGATTTGTTCTCATACCCGGC 1800  
Qy |||||  
1801 AGGCTATGACTTCTCATCGTCTATCGTCTGACTCTCTTCTTACCTATATAT 1860  
Db |||||  
1801 AGGCTATGACTTCTCATCGTCTATCGTCTGACTCTCTTCTTACCTATATAT 1860  
Qy |||||  
1861 TATTCATCCGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGG 1920  
Db |||||  
1861 TATTCATCCGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGG 1920  
Qy |||||  
1921 CGTCCGAATTC 1932  
Db |||||  
1921 CGTCCGAATTC 1932

RESULT 2  
US-09-518-386B-4  
; Sequence 4, Application US/09518386B  
; Patent No. 6365386  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yukaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/09/518/386B  
; CURRENT FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3969  
; TYPE: DNA  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (517)..(518)  
; NAME/KEY: intron  
; LOCATION: (784)..(898)  
; NAME/KEY: intron

```
; LOCATION: (1016)..(1087)
; NAME/KEY: intron
; LOCATION: (1180)..(1302)
; NAME/KEY: intron
; NAME/KEY: intron
; LOCATION: (1518)..(1600)
; NAME/KEY: intron
; LOCATION: (1635)..(1723)
; NAME/KEY: intron
; LOCATION: (1867)..(1939)
; NAME/KEY: intron
; LOCATION: (2000)..(2081)
; NAME/KEY: intron
; LOCATION: (2182)..(2257)
; NAME/KEY: intron
; LOCATION: (2355)..(2431)
; NAME/KEY: intron
; LOCATION: (2543)..(2618)
; NAME/KEY: intron
; LOCATION: (2653)..(2742)
; NAME/KEY: intron
; LOCATION: (2815)..(2962)
; NAME/KEY: intron
; LOCATION: (3051)..(3113)
; NAME/KEY: intron
; LOCATION: (3172)..(3247)
; NAME/KEY: intron
; LOCATION: (3322)..(3398)
; NAME/KEY: intron
; LOCATION: (3424)..(3513)
; NAME/KEY: polyA_site
; LOCATION: (3865)..(3866)
; NAME/KEY: intron
; LOCATION: (653)..(734)
; US-09-518-3868-4

Query Match      18.3%; Score 353.2; DB 3; Length 3969;
Best Local Similarity 99.2%; Pred. No. 3.1e-97;
Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1514 TTGGTTGGTTGGGATTTGCTGCGCGAGATGAGGCGCTTCTGTTGCTCACTCTCCG 1573
Db 3508 TCGCAGTGGTTGGGATTTGCTGCGCGAGATGAGGCGCTTCTGTTGCTCACTCTCCG 3567

Qy 1574 TCGGGTCCAGTTCGAGGCCCATCTCTCATCCAGATGAGGACATCACTTGATCAT 1633
Db 3568 TCGGGTCCAGTTCGAGGCCCATCTCTCATCCAGATGAGGACATCACTTGATCAT 3627

Qy 1634 TTCCCGTCCCTCGAATCGTTGGTAGAGAGAGGGGTACAGATGCGTTTCAGGTCAA 1693
Db 3628 TTCCCGTCCCTCGAATCGTTGGTAGAGAGAGGGGTACAGATGCGTTTCAGGTCAA 3687

Qy 1694 GCCGGTCCGAATGAGTTGATTTCTTATATGTTAAGAGAAAGTTCTATATCTCAGAAATGTGTG 1753
Db 3688 GCCGGTCCGAATGAGTTGATTTCTTATATGTTAAGAGAAAGTTCTATATCTCAGAAATGTGTG 3747

Qy 1754 ACTAGGACAATGCCTCTCTTTGATFCGATTTGTTTCTCATACCGGGGAGCGGCTATGACT 1813
Db 3748 ACTAGGACAATGCCTCTCTTTGATFCGATTTGTTTCTCATACCGGGGAGCGGCTATGACT 3807

Qy 1814 TCTACGTGCTATCGTGGCTCTGAGCTCTCTCTTACCTATATATATTTCATCCG 1871
Db 3808 TCTACGTGCTATCGTGGCTCTGAGCTCTCTCTTACCTATATATATTTCATCCG 3865
```

## RESULT 3

```
US-08-277-031B-4
; Sequence 4, Application US/08277031B
; Patent No. 6620593
; GENERAL INFORMATION:
; APPLICANT: Hayashi, Koji
; APPLICANT: Sakaki, Toshiyuki
; APPLICANT: Yabusaki, Yoshiyasu
; APPLICANT: Komai, Koichiro
```

## RESULT 4

```
US-09-023-655-1062
; Sequence 1062, Application US/09023655
```

```
; APPLICANT: Kaneko, Hideo
; APPLICANT: Nakatsuka, Iwao
; TITLE OF INVENTION: METHOD FOR SAFETY EVALUATION OF
; TITLE OF INVENTION: CHEMICAL COMPOUND USING RECOMBINANT YEAST EXPRESSING
; TITLE OF INVENTION: HUMAN CYTOCHROME P450
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5inch, 1.44MB
; COMPUTER: IBM PC
; OPERATING SYSTEM: Dos 5.0
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,031B
; FILING DATE: 19-JULY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-201120/1993
; APPLICATION NUMBER: JP-180246/1993
; APPLICATION NUMBER: JP-208279/1993
; FILING DATE: 20-07-1993
; FILING DATE: 21-07-1993
; FILING DATE: 30-07-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Raymond C. Stewart
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 20-3530P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1512
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-277-031B-4

Query Match      2.5%; Score 48.4; DB 4; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.00031;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

Qy 997 AAAAGCTGTCCGATGAGGAGTACTCGCTCAGATCAGTAACCTGTTATTTCGTGATATG 1056
Db 863 AAGCTCTGTCGATCTGAGCTCGTGGCCCAATCAATTATCTTTATTTTGTGCTATG 922

Qy 1057 AAATCTTTCGACAGCTTTGACATGGATGTTTCCCGACTCTCAGAACAAAGCGTTC 1116
Db 923 AAACACGAGCAGTGTCTCTCTCTTCATATGATGAACTGGCCACTCACCCGTATGTC 982

Qy 1117 AGGATAAACTTCGAGAAGAAATTTGTGATCGACGACGACGATATGCTACGCTAGACGAAC 1176
Db 983 ASCAGAACTTCGAGGAGAAATTTGATGAGTTTATCCCAATAAGGACCAACCCTATG 1042

Qy 1177 TTAATGCGTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTGTCTAGACC 1230
Db 1043 ATACTGTGTACAGATGGAGTATCTTGATGCTGGTGGTGAATGAAACGCTCAGATTATCC 1102

Qy 1231 CTCCTAGTCCGTATCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
Db 1103 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGAGATCA 1150
```

Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
FILE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1062:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2059 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: G181373  
US-09-023-655-1062

Query Match 2.5%; Score 48.4; DB 4; Length 2059;  
Best Local Similarity 50.7%; Pred. No. 0.00038;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;  
QY 997 AAAAGCTGTCGATGAGGAGTACTCGCTGAGATCAGTAACTGTTATTTGCTGGATG 1056  
DB 953 AAGCTCTGTCGATGAGGAGTCTGCGCCCAATCAATATCTTTATTTTCTGGCTATG 1012  
QY 1057 AAATCTTCGACAGTCTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTTTC 1116  
DB 1013 AAACACGAGCAGTGTCTCTCTCTTATGATGATGAGCTGGCCACTCACCCTGATGTC 1072  
QY 1117 AGGATAAACTTCGAGAGAAATTTGTGATGATGATGATGATGATGATGATGATGATG 1176  
DB 1073 AGCAGAACTCGAGAGGAAATTTGATGATGATGATGATGATGATGATGATGATGATG 1132  
QY 1177 TTAATGCGTTCG-----CTTATCTCGAAGCGTTTGTAAAGGATGCTCTGCTAGAC 1230  
DB 1133 ATCTGTCTACAGTGGAGTATCTTGCATGTTGTTGATGATGATGATGATGATGATGATG 1192  
QY 1231 CTCCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAGACTTCA 1278  
DB 1193 CAATTGCTGATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGATGATCA 1240

RESULT 5  
US-09-144-367-1

Sequence 1, Application US/09144367  
Patent No. 6432639  
GENERAL INFORMATION:  
APPLICANT: Lichter, Jay  
APPLICANT: Guido, Marco  
TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
FILE OF INVENTION: SEQ-12P  
CURRENT APPLICATION NUMBER: US/09/144,367  
CURRENT FILING DATE: 1998-08-31  
PRIOR APPLICATION NUMBER: 60/058,612  
PRIOR FILING DATE: 1997-09-10  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2759  
TYPE: DNA  
ORGANISM: H. sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (70)...(1581)  
OTHER INFORMATION: Human CYP3A4 cDNA reference sequence  
US-09-144-367-1

Query Match 2.5%; Score 48.4; DB 4; Length 2759;  
Best Local Similarity 50.7%; Pred. No. 0.00046;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;  
QY 997 AAAAGCTGTCGATGAGGAGTACTCGCTGAGATCAGTAACTGTTATTTGCTGGATG 1056  
DB 932 AAGCTCTGTCGATGAGGAGTCTGCGCCCAATCAATATCTTTATTTTCTGGCTATG 991  
QY 1057 AAATCTTCGACAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1116  
DB 992 AAACACGAGCAGTGTCTCTCTTATGATGATGATGATGATGATGATGATGATGATG 1051  
QY 1117 AGGATAAACTTCGAGAGAAATTTGTGATGATGATGATGATGATGATGATGATGATG 1176  
DB 1052 AGCAGAACTCGAGAGGAAATTTGATGATGATGATGATGATGATGATGATGATGATG 1111  
QY 1177 TTAATGCGTTCG-----CTTATCTCGAAGCGTTTGTAAAGGATGCTCTGCTAGAC 1230  
DB 1112 ATCTGTCTACAGTGGAGTATCTTGCATGTTGTTGATGATGATGATGATGATGATGAT 1171  
QY 1231 CTCCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAGACTTCA 1278  
DB 1172 CAATTGCTGATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGATGATCA 1219

RESULT 6  
US-09-716-129-12  
Sequence 12, Application US/09716129  
Patent No. 6632920  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 36 Human Secreted Proteins  
FILE OF INVENTION: P2025P1  
CURRENT APPLICATION NUMBER: US/09/716,129  
CURRENT FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/076,053  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/076,057  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/076,052  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/076,054  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/076,051  
NUMBER OF SEQ ID NOS: 186  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 12  
LENGTH: 2160  
TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-716-129-12

Query Match      2.5%; Score 47.4; DB 4; Length 2160;
Best Local Similarity 70.8%; Pred. No. 0.0008;
Matches 63; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1826 ATCCGCTCTGGACTCTCTTACCTATATATTCATCCGAAAAA 1885
DB 2069 ATTGTTCCGTTGACTAAGAACTTGACCTAAATAATCCCAAGTATAAAAAA 2128

QY 1886 AAAAAAAGGCGCGCTCGAG 1914
DB 2129 AAAAAAAGGCGCGCTCGG 2157

RESULT 7
US-09-148-545-106.
; Sequence 106, Application US/09148545
; Patent No. 6590075
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/1487545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
```



EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,674  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,669  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 05-Sep-1997  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
NUMBER OF SEQ ID NOS: 280  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 105  
LENGTH: 1822

Query Match 2.4%; Score 46.8; DB 4; Length 1822;  
Best Local Similarity 85.0%; Pred. No. 0.0011;

Matches 51; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1854 TATATATTATTCCTCCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCTCGA 1913

DB 1755 TAAATATATTTTGTACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGSGCGCTCGA 1814

```

; GENERAL INFORMATION:
; APPLICANT: Siminsky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1747
US-08-948-564-15

Query Match 2.3%; Score 45; DB 3; Length 1831;
Best Local Similarity 52.8%; Pred. No. 0.0039;
Matches 121; Conservative 0; Mismatches 105; Indels 3; Gaps 1;

QY 1008 GATGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTGGATATGAACTTCTTCG 1067
Db 1037 GATCGTCAGTTGAGGATGATTAAATGACAATGCTTATTGCGGTCATGAACAACGGCT 1096
QY 1068 ACAGTCTTGATGATGATGTTTCCCGACTCTCAGAGACAAACCGTTTCAGGTAACCTT 1127
Db 1097 GCAGTTCTTACTTGGGCAGTTTTCCTCCTAGCTCAAAATCCTAGCAAAATGAAGAAGGCT 1156
QY 1128 CGAGAAGA--AAATTCTCAGATCGACACGATATGCTTACGCTAGACGAACCTTAATGCG 1184
Db 1157 CAACAGAGGTAGATTGTTGCTGGTACGGGAGGCCAACTTTTGAATCACTTAGGAA 1216
QY 1185 TTGCTTATCTCGAAGCGTTTGTAGAGTCTCTTCGCTAGACCCCTC 1233
Db 1217 TTGAGTACATTAAGATTGATTGTTGTGGAGGCTCTTCGTTTATACCCCTC 1265

RESULT 10
US-09-709-103-3
; Sequence 3, Application US/09709103
; Patent No. 6733991

; GENERAL INFORMATION:
; APPLICANT: Cismowski, Mary
; APPLICANT: Duzic, Emir
; TITLE OF INVENTION: AGS Proteins and Nucleic Acid Molecules and Uses Therefor
; FILE REFERENCE: 60388-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/709,103
; CURRENT FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(996)
; OTHER INFORMATION:
US-09-709-103-3

Query Match 2.3%; Score 44.8; DB 4; Length 1801;
Best Local Similarity 95.8%; Pred. No. 0.0044;
Matches 46; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1872 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCGCTCGAGCGGC 1919
Db 1719 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCGCTCGAGCATGC 1766

RESULT 11
US-09-439-410A-3
; Sequence 3, Application US/09439410A
; Patent No. 6746852
; GENERAL INFORMATION:
; APPLICANT: Cismowski, Mary
; APPLICANT: Duzic, Emir
; TITLE OF INVENTION: AGS PROTEIN AND NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: 1919/60388-B
; CURRENT APPLICATION NUMBER: US/09/439,410A
; CURRENT FILING DATE: 1999-11-11
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(996)
; OTHER INFORMATION:
US-09-439-410A-3

Query Match 2.3%; Score 44.8; DB 4; Length 1801;
Best Local Similarity 95.8%; Pred. No. 0.0044;
Matches 46; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1872 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCGCTCGAGCGGC 1919
Db 1719 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGCGCGCTCGAGCATGC 1766

RESULT 12
US-09-221-013A-7
; Sequence 7, Application US/09221013A
; Patent No. 6495740
; GENERAL INFORMATION:
; APPLICANT: Arioli, Antonio
; APPLICANT: Williamson, Richard E.
; APPLICANT: Betzner, Andreas S.
; APPLICANT: Peng, Liangcai
; TITLE OF INVENTION: Manipulation of cellulose and/or Beta-1,4-glucan
; FILE REFERENCE: 96-98
; CURRENT APPLICATION NUMBER: US/09/221,013A
; CURRENT FILING DATE: 1998-12-23
```



```
; FILE REFERENCE: 0919
; CURRENT APPLICATION NUMBER: US/09/372,422A
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: US 60/098,692
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1333
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (142)...(1018)
US-09-372-422A-9
```

```
Query Match 2.3%; Score 44; DB 4; Length 3828;
Best Local Similarity 83.3%; Pred. No. 0.013;
Matches 50; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1854 TATATATTATCCATCCGAAAGAAAAAAGAAAAAAGAAAAAAGCGCGCTCGA 1913
Db 3746 TACACACTTGTACAAAAAAGAAAAAAGAAAAAAGCGCGCTCGA 3805
```

```
RESULT 13
US-09-144-367-8
; Sequence 8, Application US/09144367
; Patent No. 6432639
; GENERAL INFORMATION:
; APPLICANT: Lichter, Jay
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/09/144,367
; CURRENT FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/058,612
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 352
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: Other
; LOCATION: (0)...(0)
US-09-144-367-8
```

```
Query Match 2.3%; Score 43.6; DB 4; Length 352;
Best Local Similarity 57.2%; Pred. No. 0.0036;
Matches 79; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 1002 CTGTCCGATGAGGAGTACTCGCTCAGATCAGTAACCTGTATTTCGTGGATATGAACT 1061
Db 123 CTGTCCGATGAGGAGTACTCGCTCAGATCAGTAACCTGTATTTCGTGGATATGAACT 182

Qy 1062 TCTTCGACATCTTGACATGGATGTTTCACCGACTTCAGAGACAAAGCCGTTACGAT 1121
Db 183 ACCGACGAGTCTTCTCTCTTCAATGATGAAGTGGCCACTCACCCCTGATGTCCAGCAG 242

Qy 1122 AAACCTCGAGAGAAATT 1139
Db 243 AAACCTCGAGAGAAATT 260
```

```
RESULT 14
US-09-372-422A-9
; Sequence 9, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
; APPLICANT: Rudolf Jung
; APPLICANT: Francois Barrieu
; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
```

```
; FILE REFERENCE: 0919
; CURRENT APPLICATION NUMBER: US/09/372,422A
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: US 60/098,692
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1333
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (142)...(1018)
US-09-372-422A-9
```

```
Query Match 2.2%; Score 42.8; DB 3; Length 1333;
Best Local Similarity 87.0%; Pred. No. 0.015;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1855 ATATATTATTCATCCGAAAAAAGAAAAAAGAAAAAAGCGCGCG 1908
Db 1280 ATCGATTATTCCTCCCAAAAAAAGAAAAAAGAAAAAAGCGCGCG 1333
```

```
RESULT 15
US-09-370-807-7
; Sequence 7, Application US/09370807
; Patent No. 6297034
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Falco, S. Carl
; APPLICANT: Rafalski, J. Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: N-End Rule Pathway Enzymes
; FILE REFERENCE: BB-1199
; CURRENT APPLICATION NUMBER: US/09/370,807
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: 60/096,225
; EARLIER FILING DATE: August 12, 1998
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2407
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-370-807-7
```

```
Query Match 2.2%; Score 42.8; DB 3; Length 2407;
Best Local Similarity 80.6%; Pred. No. 0.022;
Matches 50; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1840 CTCCTCTTCTTACCTATATATATTCATCCGAAAAAAGAAAAAAGAAAAAAGAAAAA 1899
Db 1939 CTATCATCTTAGCGTTACATTCGCCATGAAAAAAGAAAAAAGAAAAAAGAAAAA 1998

Qy 1900 AA 1901
Db 1999 AA 2000
```

Search completed: December 14, 2004, 14:35:33  
Job time : 161.79 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2004, 19:45:42 ; Search time 893.808 Seconds  
(without alignments)  
11346.825 Million cell updates/sec

Title: US-10-066-007A-2

Perfect score: 1932

Sequence: 1 gaattcgacagagccacc.....agccggctcgtgcgaattc 1932

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_23Sep04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002as:\*
- 7: Geneseq2002bs:\*
- 8: Geneseq2003as:\*
- 9: Geneseq2003bs:\*
- 10: Geneseq2003cs:\*
- 11: Geneseq2003ds:\*
- 12: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1932	100.0	1932	3	AAA64472
2	353.2	18.3	3969	3	AAA64473
3	51.8	2.7	2000	8	ADA71938
4	49.4	2.6	1501	6	ABL89830
5	48.4	2.5	991	6	ABK97688
6	48.4	2.5	1062	6	ABK97694
7	48.4	2.5	1365	6	ABK97693
8	48.4	2.5	1458	12	ADJ87520
9	48.4	2.5	1458	12	ADJ87520
10	48.4	2.5	1512	2	AAQ87717
11	48.4	2.5	1512	2	AAT17399
12	48.4	2.5	1512	2	AAT28383
13	48.4	2.5	1512	6	ABK97692
14	48.4	2.5	1512	9	ACF06054
15	48.4	2.5	1512	12	ADL18671
16	48.4	2.5	2011	6	ABN95088
17	48.4	2.5	2059	11	ADJ17136
18	48.4	2.5	2759	2	AAK28295
19	48.4	2.5	2764	10	ABX77172
20	48.4	2.5	2768	10	ADG87158
21	48.4	2.5	2768	10	ADG89415

22 48.4 2.5 2768 10 ADG89349  
23 48.4 2.5 2849 6 ABN95612  
24 48.2 2.5 312 8 ABX39446  
25 48.2 2.5 1645 3 AAC77026  
26 47.6 2.5 3059 3 AAC99040  
27 47.4 2.5 1410 12 ADK14883  
28 47.4 2.5 2160 2 ADZ10641  
29 47 2.4 217 3 ADF56843  
30 47 2.4 445 5 ABV39053  
31 46.8 2.4 1712 2 AAV34249  
32 46.8 2.4 1712 8 ACD08120  
33 46.8 2.4 1822 2 AAV34248  
34 46.8 2.4 1822 8 ACD08119  
35 46.8 2.4 6004 10 ADB58866  
36 46.8 2.4 6004 10 ADP65872  
37 46.4 2.4 1552 4 ABL14451  
38 46 2.4 2190 12 ADH13718  
39 45.4 2.3 576 5 ABV44839  
40 45.4 2.3 5002 12 ADQ23512  
41 45.2 2.3 413 8 ABX36570  
42 45.2 2.3 1608 6 ABK97643  
43 45.2 2.3 1608 10 ADD29816  
44 45.2 2.3 1660 12 ADL90278  
45 45.2 2.3 1971 6 ABL62314

## ALIGNMENTS

RESULT 1  
AAA64472  
ID AAA64472 standard; cDNA; 1932 BP.  
XX  
AC AAA64472;  
XX  
DT 15-SEP-2003 (revised)  
DT 02-JAN-2001 (first entry)  
XX  
DE cDNA encoding an astaxanthin synthetase polypeptide.  
XX  
KW Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;  
KW antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.  
XX  
OS Xanthophyllomyces dendrorhous.  
XX  
FH Key Location/Qualifiers  
FT CDS 33..1706  
FT /\*tag= a  
FT /product= "astaxanthin synthetase"  
FT polyA\_signal 1871  
FT /\*tag= b

XX  
PN BP1035206-A1.  
XX  
PD 13-SEP-2000.  
XX  
PF 03-MAR-2000; 2000EP-00104430.  
XX  
PR 09-MAR-1999; 99EP-00104668.  
PR 01-FEB-2000; 2000EP-00101666.  
XX  
(HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
PA Aq87717 Human cyt  
XX  
PI Aat17399 Human der  
XX  
XX Aat28383 Human cyt  
XX  
XX Abk97692 DNA encod  
XX  
XX Acf06054 Human cyt  
XX  
XX Adl18671 Human cyt  
XX  
XX Abn95088 Gere #158  
XX  
XX Adj17136 Human cDN  
XX  
XX Aax28295 Human cyt  
XX  
XX Abx77172 DNA seque  
XX  
XX Adg87158 Human CAR  
XX  
XX Adg89415 Cancer de

Novel polynucleotide encoding astaxanthin synthase useful for producing recombinant cells for producing astaxanthin from beta-carotene.

Claim 3; Page 24-27; 46pp; English.

CC The present sequence encodes an astaxanthin synthetase polypeptide of  
 CC *Phaffia rhodozyma*. The enzyme is involved in the last step of the  
 CC astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin. P.  
 CC *rhodozyma* is a carotenogenic yeast strain. The astaxanthin synthetase  
 CC polynucleotides and polypeptides are useful for producing astaxanthin.  
 CC Astaxanthin is an antioxidant which may be used to protect living cells  
 CC against diseases such as cancer. Astaxanthin is also used as a colouring  
 CC reagent, e.g. in farmed fish like salmon to impart an orange-red  
 CC coloration. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1932 BP; 495 A; 456 C; 481 G; 500 T; 0 U; 0 Other;

Query March 100.0%; Score 1932; DB 3; Length 1932;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGCGCAGGCGCCACTTCTTCCCATATGTTTCATCTTGGTCTTGTCTCACAGGTG 60  
 DB 1 GAATTCGGCAGGCGCCACTTCTTCCCATATGTTTCATCTTGGTCTTGTCTCACAGGTG 60  
 QY 61 CTTTAGGCTGGCTGCTTTCTCATGGGCATCCATAGCGTTCTTTCAGTCTTTTACCTCGCTC 120  
 DB 61 CTTTAGGCTGGCTGCTTTCTCATGGGCATCCATAGCGTTCTTTCAGTCTTTTACCTCGCTC 120  
 QY 121 CGAGGGCATCTTCACTGTATACCTTCAGGCGCCGGAATCAACCACTTCTTACAGGCA 180  
 DB 121 CGAGGGCATCTTCACTGTATACCTTCAGGCGCCGGAATCAACCACTTCTTACAGGCA 180  
 QY 181 ATTTTATAGCATCTCTCAGCTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAT 240  
 DB 181 ATTTTATAGCATCTCTCAGCTCGTACAGGTGAAGAGCATGCGAAGTACAGAGAAAT 240  
 QY 241 ACGGAAGCACTCCGGTTGCTGGGATCGCTGGAGCAGCAGCCGCTTGAATCCGACCGATC 300  
 DB 241 ACGGAAGCACTCCGGTTGCTGGGATCGCTGGAGCAGCAGCCGCTTGAATCCGACCGATC 300  
 QY 301 CGAAGTCTTCAACCATGTGATGAAGAAGCCTACGACTATCCGAAACCTGGTATGSCCG 360  
 DB 301 CGAAGTCTTCAACCATGTGATGAAGAAGCCTACGACTATCCGAAACCTGGTATGSCCG 360  
 QY 361 CTCGAGTCTCAGATTCGTAACCGAGATGCTGTGTACGCGGAGAGGTGAGACTCATA 420  
 DB 361 CTCGAGTCTCAGATTCGTAACCGAGATGCTGTGTACGCGGAGAGGTGAGACTCATA 420  
 QY 421 AGCGACATCGAAGGATCATGATCCCTCTCTGTCCGCTCAGGCGCGTTAAGTCGATGCTC 480  
 DB 421 AGCGACATCGAAGGATCATGATCCCTCTCTGTCCGCTCAGGCGCGTTAAGTCGATGCTC 480  
 QY 481 CAATTTCTTGAAGAAAGGTATGGAATGTCGACAGATGATCGAGGATGCGGCTGAGA 540  
 DB 481 CAATTTCTTGAAGAAAGGTATGGAATGTCGACAGATGATGGAAGGTGCGGCTGAGA 540  
 QY 541 AGGATATGCGCTGGGAGAGTCCGCGTGAAGAAAGGCAACAGAGACTCGAGACCGAAG 600  
 DB 541 AGGATATGCGCTGGGAGAGTCCGCGTGAAGAAAGGCAACAGAGACTCGAGACCGAAG 600  
 QY 601 GAGTCGATGAAGGATGGGTCGCTGAGCTACTCTGGAGCTATCGCTCTTGCAGGAT 660  
 DB 601 GAGTCGATGAAGGATGGGTCGCTGAGCTACTCTGGAGCTATCGCTCTTGCAGGAT 660  
 QY 661 TTGACTATAAGAGCGACTCGCTCCAGAAACAAGACCAATGAGCTCTATGTGCTTTTGTG 720  
 DB 661 TTGACTATAAGAGCGACTCGCTCCAGAAACAAGACCAATGAGCTCTATGTGCTTTTGTG 720  
 QY 721 GACTTACCGATGGGTTGCTCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTG 780  
 DB 721 GACTTACCGATGGGTTGCTCTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTG 780  
 QY 781 TACCTTACTTCCGAATATGAACCGAGACATGAGATACCTTTGACTCAAGGATTAGCAG 840  
 DB 781 TACCTTACTTCCGAATATGAACCGAGACATGAGATACCTTTGACTCAAGGATTAGCAG 840  
 QY 841 TTTCCCGCAGAGTTGGGATCGAGCTTATGAGAGCAAGAGAGCGCGCTGCTTGGCTCAG 900

DB 841 TTTCCCGCAGAGTTGGGATCGAGCTTATGAGCAAAAGAGCGCGTCTTGGCTCAG 900  
 QY 901 CTTCCGATCAGCTGTTGATAAAAGAGATGTTCAAGTCCGGATATCCTAAGTCTCCTAG 960  
 DB 901 CTTCCGATCAGCTGTTGATAAAAGAGATGTTCAAGTCCGGATATCCTAAGTCTCCTAG 960  
 QY 961 TGAGAGCAAAACATCGCGCAACCTGCTGAATCTCAAAAGCTGTCGATGAGAGGTAC 1020  
 DB 961 TGAGAGCAAAACATCGCGCAACCTGCTGAATCTCAAAAGCTGTCGATGAGAGGTAC 1020  
 QY 1021 TCGCTCAGATCAGTAACTGTTATTTCTGGATATGAAACTTCTTCGACAGTCTTGACAT 1080  
 DB 1021 TCGCTCAGATCAGTAACTGTTATTTCTGGATATGAAACTTCTTCGACAGTCTTGACAT 1080  
 QY 1081 GGAATGTTTACCGACTCTCAGAGACAAAGCGTTTCCAGGATAAACTTCGAGAGAAATTT 1140  
 DB 1081 GGAATGTTTACCGACTCTCAGAGACAAAGCGTTTCCAGGATAAACTTCGAGAGAAATTT 1140  
 QY 1141 GTCAGATCGACACGAGATATGCTACGCTAGACGAACTTAATGCGTTGCTTATCTGAAAG 1200  
 DB 1141 GTCAGATCGACACGAGATATGCTACGCTAGACGAACTTAATGCGTTGCTTATCTGAAAG 1200  
 QY 1201 CGTTTGTTAAGAGTCTCTTCGCTAGACCCCTCTAGTCCGATGCTTAACCGTGAATGCT 1260  
 DB 1201 CGTTTGTTAAGAGTCTCTTCGCTAGACCCCTCTAGTCCGATGCTTAACCGTGAATGCT 1260  
 QY 1261 TAAAGGATGAAGACTTCAATCCACTTCCGAGCTCTCATTTGCTCGAGATGGGTCCGTC 1320  
 DB 1261 TAAAGGATGAAGACTTCAATCCACTTCCGAGCTCTCATTTGCTCGAGATGGGTCCGTC 1320  
 QY 1321 TCAACGAGTCCCGATCAGGAAAGGACGATGCTCATGCTTCCGTTGTTCAACATCAATC 1380  
 DB 1321 TCAACGAGTCCCGATCAGGAAAGGACGATGCTCATGCTTCCGTTGTTCAACATCAATC 1380  
 QY 1381 GTTCAAAAGTTCATTTATGGAGAGATGCAAGAAATTCAGACCGGAGAGTGGCTTGAG 1440  
 DB 1381 GTTCAAAAGTTCATTTATGGAGAGATGCAAGAAATTCAGACCGGAGAGTGGCTTGAG 1440  
 QY 1441 ACGTAAACAGACTCGCTCAACAGTATTAAGACCCCTATGACACCGAGGCTTTATCT 1500  
 DB 1441 ACGTAAACAGACTCGCTCAACAGTATTAAGACCCCTATGACACCGAGGCTTTATCT 1500  
 QY 1501 CTGACCCAGAGCTTGTGTTGGTGGGATTTGCTGTGCGCGAGATGAAGGCTTCTTGT 1560  
 DB 1501 CTGACCCAGAGCTTGTGTTGGTGGGATTTGCTGTGCGCGAGATGAAGGCTTCTTGT 1560  
 QY 1561 TTGTCACTCTCCGTCGGGTCCAGTTCGAGCCCATCATCTCTCATCCAGAGTACGAGCACA 1620  
 DB 1561 TTGTCACTCTCCGTCGGGTCCAGTTCGAGCCCATCATCTCTCATCCAGAGTACGAGCACA 1620  
 QY 1621 TCACCTTGATCATTTCCGCTCCTCGAATCGTTGGTAGAGAGAGGAGGATCCAGATGC 1680  
 DB 1621 TCACCTTGATCATTTCCGCTCCTCGAATCGTTGGTAGAGAGAGGAGGATCCAGATGC 1680  
 QY 1681 GTTTGCGAGTCAAGCGCTCGAATGAGTGTTCATATGTTAAGAGAGTTCATAT 1740  
 DB 1681 GTTTGCGAGTCAAGCGCTCGAATGAGTGTTCATATGTTAAGAGAGTTCATAT 1740  
 QY 1741 CTGAGAAATGTGACTAGGACAAATGCTTCTTTGATCGATTTGTTTCTCATACCGGCG 1800  
 DB 1741 CTGAGAAATGTGACTAGGACAAATGCTTCTTTGATCGATTTGTTTCTCATACCGGCG 1800  
 QY 1801 AGGCGCTATGACTTCTACGCTCTCTCGCTCTGAGCTCTCTTCTTACCCCTATAT 1860  
 DB 1801 AGGCGCTATGACTTCTACGCTCTCTCGCTCTGAGCTCTCTTCTTACCCCTATAT 1860  
 QY 1861 TATTCCATCCGAAATTTTCAATGAGTGTTCATATGTTAAGAGAGTTCATAT 1920  
 DB 1861 TATTCCATCCGAAATTTTCAATGAGTGTTCATATGTTAAGAGAGTTCATAT 1920  
 QY 1921 CGTCCGAAATTC 1932

[illegible]

DR P-PSDB; AAB08713.  
XX Novel polynucleotide encoding astaxanthin synthase useful for producing  
PT recombinant cells for producing astaxanthin from beta-carotene.  
XX  
XX Claim 4; Page 30-33; 46pp; English.  
XX  
XX The present sequence encodes an astaxanthin synthetase polypeptide of  
CC Phaffia rhodozyma. The enzyme is involved in the last step of the  
CC astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin. P.  
CC rhodozyma is a carotenogenic yeast strain. The astaxanthin synthetase  
CC polynucleotides and polypeptides are useful for producing astaxanthin.  
CC Astaxanthin is an antioxidant which may be used to protect living cells  
CC against diseases such as cancer. Astaxanthin is also used as a colouring  
CC reagent, e.g. in farmed fish like salmon to impart an orange-red  
XX coloration. (Updated on 15-SEP-2003 to standardise OS field)  
XX  
XX Sequence 3969 BP; 937 A; 942 C; 869 G; 1221 T; 0 U; 0 Other;  
SQ  
Query Match 18.3%; Score 353.2; DB 3; Length 3969;  
Best Local Similarity 99.2%; Pred. No. 4.7e-87;  
Matches 355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1514 TTGCTTTGGTGGGATTTGCTGCGCGAGATGAAGGCTTCTTTGTTGCTACTCTCCG 1573  
DB 3508 TCGCAGTGGTGGGATTTGCTGCGCGAGATGAAGGCTTCTTTGTTGCTACTCTCCG 3567  
QY 1574 TCGGCTTCAGTTCGAGGCCATCATCTCTCATCCAGAGTACGAGACATCACTTGATCAT 1633  
DB 3568 TCGGCTTCAGTTCGAGGCCATCATCTCTCATCCAGAGTACGAGACATCACTTGATCAT 3627  
QY 1634 TTCCCGTCTCGATCGTTGGTAGAGAGAGGCGGTACCAGATCGGTTTGCAAGTCAA 1693  
DB 3628 TTCCCGTCTCGATCGTTGGTAGAGAGAGGCGGTACCAGATCGGTTTGCAAGTCAA 3687  
QY 1694 GCCGGTCAATGAGTTGATTTCTTCATATGTTTAAGAGAGTTCTATATCTCAGAAATGTTG 1753  
DB 3688 GCCGGTCAATGAGTTGATTTCTTCATATGTTTAAGAGAGTTCTATATCTCAGAAATGTTG 3747  
QY 1754 ACTAGCAATGCTTCTTTGATCGATTTGTTTCTCATACCGGCGAGCGCTGATGACT 1813  
DB 3748 ACTAGCAATGCTTCTTTGATCGATTTGTTTCTCATACCGGCGAGCGCTGATGACT 3807  
QY 1814 TCTAGTCTGTATCGTCTGCTCTGACTCTCTTCTTACCCTATATATTTCCATCCG 1871  
DB 3808 TCTAGTCTGTATCGTCTGCTCTGACTCTCTTCTTACCCTATATATTTCCATCCG 3865  
RESULT 3  
ID ADA71938  
XX ADA71938 standard; DNA; 2000 BP.  
XX  
XX AC ADA71938;  
XX  
XX DT 20-NOV-2003 (first entry)  
XX  
XX DE Rice gene, SEQ ID 5263.  
XX  
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene; ds.  
XX  
XX OS Oryza sativa.  
XX  
XX PN WO2003000898-A1.  
XX  
XX PD 03-JAN-2003.  
XX  
XX PF 22-JUN-2001; 2001WO-IB001105.  
XX  
XX PR 22-JUN-2001; 2001WO-IB001105.  
XX  
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC the expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
SQ  
Query Match 2.7%; Score 51.8; DB 8; Length 2000;  
Best Local Similarity 9.8%; Pred. No. 0.0021;  
Matches 72; Conservative 343; Mismatches 310; Indels 10; Gaps 3;  
QY 421 AGCGACATCGAAGATCATGATCCCTCTCTGTCGCTCAGCGCGTAAAGTCGATGCC 480  
DB 48 RKCSGCGKCMTRKSKWYSASSASGRTGSKWSGSYSGKMKRYKRKRWRGRRG 107  
QY 481 CAATTTCTTAGAAAAGCTATGAACTGTGCAACAGATGATGAGATCGGCTGAGA 540  
DB 108 MRSRMWNRVRCARSRVAGSGRNMGGKSRMSYMWVCYARGCSCKKKS--KGGS 165  
QY 541 AGGATATCGCGTGGAGAGTCGCGCGGTGAAAGAGAGCAACAGACTCGAGACCGAAG 600  
DB 166 WGTCTCRGARGSGWSGSAKYKSGMSKRMWMSGCGRRSAYRYGTSRYGYK 225  
QY 601 GAGTCGATGAAGGATTTGGTTCGCTCGAGTACTCTCGAGCTCATGCTCTTCAGGAT 660  
DB 226 KMTYSAGRCMPAYMTTSYWCSSYTWCRKRKRSMWMMKRWKRSRYSYMSYKMW 285  
QY 661 TTGACTATAAGAGCAGCTCGCTCCAGAACAGACCAATGAGCTCTATCCGCTTTGTCG 720  
DB 286 MCTAYKSYSRWCYMYRGGWGRGATRYWGRGYNSRMAMMYKMYRYGYKMGKRWAG 345  
QY 721 GACTTACGATGGGTTTGTCTTACTTGGACTCGTCAAGGCTATCATGTGGGATTTTG 780  
DB 346 RMMRSMCRKWSKACYMYRWRVMTTRRRRWAKSSRTSRKRWKRWKRWKRWKRWKRW 405  
QY 781 TACCTTACTTCGAACTATGAACGAGACATGAGA-----TACCTTTGACTCAAGAT 834  
DB 406 MRSCKEARWMECRSGRAWKNGCRGMCCKMSYGMWVKWKSWMKRWKRWKRWKRWKRW 465  
QY 835 TAGCAGTTTCCGACGAGTTGGATCGAGCTTATGGAGCAAGAGAGCGCGCTGCTTG 894  
DB 466 KKCSTRTTMMGTGGWMTGRCRYKSGMKRKRWRWRWRWRWRWRWRWRWRWRWRWR 525  
QY 895 GTCAGCTTCCGATCAGGCTGTTGATAAAGAGATGTTCAAGGTCGGGATATCTTAAGTC 954  
DB 526 YCARKKYSISAARKACRWYRGKYWAGWMMKRYKRYKRYKRYKRYKRYKRYKRYKRY 585  
QY 955 TCTTAGTAGAGCAAAACATCGCC--GCCAACCTGCTGATCTCAAAAGCTGTCGATGA 1012  
DB 586 YYASCMKSAKAGAKMKCKSKMSAWSKMSRSRCKCKCASKSKAKRYMMGMGTSGSRM 645  
QY 1013 GGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTCGATATGAAATCTTTCACAGT 1072  
DB 646 SRWKSITCYWRKMGWSKSTCTWMTYKSYKRYKRYKRYKRYKRYKRYKRYKRYKRY 705

QY 1073 CTTGACATGATGTTTACCGACTCTCAGAGACAAAGCCGTTTCAGGATAAATTCGAGA 1132  
 Db 706 MTTAAWYTSRMTAMTGKYSGRYWTWYKICKSWKYRSWYTWYWSWAKTWKOWRYA 765  
 QY 1133 AGAAATTTGTCAGAT 1147  
 Db 766 TMMWMMYRYSWKWY 780

RESULT 4  
 ABL89830  
 ID ABL89830 standard; cDNA; 1501 BP.  
 AC ABL89830;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 392.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US016450.  
 XX  
 PR 19-MAY-2000; 2000US-0205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-122018/16.  
 DR P-PSDB; ABB89421.  
 XX

Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.

Claim 4; SEQ ID NO 392; 2081pp + Sequence Listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB9040-ABB9044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1501 BP; 375 A; 389 C; 436 G; 296 T; 0 U; 5 Other;

Query Match 2.6%; Score 49.4; DB 6; Length 1501;  
 Best Local Similarity 63.2%; Pred. No. 0.0086;  
 Matches 74; Conservative 1; Mismatches 42; Indels 0; Gaps 0;

QY 1785 TTCTCATACCGGCGGCGGCTATGACTTCTACGTCGTCTATCGTCGCTCGGACTCTC 1844  
 Db 1370 TTTCCTGSCCTGACTTCTCTTTGTCACAGACTGCATAAGTGTGACGCTTGACTATC 1429  
 QY 1845 TTCTTACCCCTATATATTATTCATCGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1901  
 Db 1430 TTITGATAAAGATTGTGATTTAAACAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1486

RESULT 5  
 ABK97688  
 ID ABK97688 standard; DNA; 991 BP.  
 AC ABK97688;  
 XX  
 DT 07-OCT-2002 (first entry)  
 XX  
 DE DNA encoding cytochrome P450 3A4 (CYP3A4) exon 7-13 protein.  
 XX  
 KW Cytochrome P450; CYP3A4; CYP3A2; CYP3A3; CYP3A4; CYP3A5; CYP3A7;  
 KW drug metabolism; drug design; drug screening; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200244213-A1.  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 28-NOV-2001; 2001WO-SE002631.  
 XX  
 PR 28-NOV-2000; 2000SE-00004366.  
 PR 11-JUN-2001; 2001SE-00002061.  
 XX  
 PA (ZAPH/) ZAPHIROPOULOS P G.  
 PA (FINT/) FINTA C.  
 XX  
 PI Zaphiropoulos PG, Finta C;  
 XX  
 DR WPI; 2002-557532/59.  
 DR P-PSDB; ABG68749.  
 XX

Novel cytochrome P450 protein in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, useful as medicament, and in evaluating drug metabolism, in drug design and drug screening.

Claim 9; Page 97-98; 131pp; English.

The invention describes a cytochrome P450 protein (I) in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, as well as sub fragments, variants and multiples of (I) having essentially the same characteristics. (I) is useful as a medicament, and for evaluating drug metabolism, in drug design, and drug screening, and in tests for adjusting the dose of drugs. This sequence encodes a novel cytochrome P450 protein

Sequence 991 BP; 302 A; 221 C; 202 G; 266 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 6; Length 991;  
 Best Local Similarity 50.7%; Pred. No. 0.013;  
 Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATGAGGAGGCTACTCGCTCAGATCAGTAACCTGTTATTTCTGATATG 1056  
 Db 342 AAGCTCTCTCGATCTGGAGCTCGTGCCCAATCAATTATCTTTATTTTCTGCTATG 401  
 QY 1057 AAATCTTCTTCACAGTCTTGACATGGATGTTTACCAGACTCTCAGAGACAAAGCGGTC 1116  
 Db 402 AAACCCAGCAGCAGTGTTCTCTCTCTTATGATGATGAGTCCACTCACCTGATGTC 461

QY 1117 AGGATAAACTTCGAGAGAAATTTGTGATCGACAGGATATGCTTACGTCAGCAAC 1176  
 Db 462 AGCAGAAATCTCAGGAGGAAATTGATGCAAGTTTACCCCAATAGGCACCCACCTATG 521

QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTCGCTAGACC 1230  
Db 522 ATACTGTCTACAGATGGAGTATCTTGACATGGTGTGATGAACGCTCAGATTATTC 581  
QY 1231 CTCCTAGTCGATGCTTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
Db 582 CAATTGCTATGAGACTTGAAGGGTCTGCAAAAAGATGTTGATCA 629

RESULT 6  
ABK97694  
ID ABK97694 standard; DNA; 1062 BP.  
XX  
AC ABK97694;  
XX  
DT 07-OCT-2002 (first entry)  
XX  
DE DNA encoding cytochrome P450 CYP3A43 exon 1-CYP3A4 exon 7-13.  
XX  
KW Cytochrome P450; CYP3A21; CYP3AP2; CYP3A43; CYP3A4; CYP3A5; CYP3A7;  
KW drug metabolism; drug design; drug screening; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200244213-A1.  
XX  
PD 06-JUN-2002.  
XX  
PF 28-NOV-2001; 2001WO-SE002631.  
XX  
PR 28-NOV-2000; 2000SE-00004366.  
PR 11-JUN-2001; 2001SE-00002061.  
XX  
XX (ZAPH/) ZAPHIROPOULOS P G.  
PA (FINT/) FINTA C.  
XX  
PI Zaphiropoulos PG, Finta C;  
XX  
XX WPI; 2002-557532/59.  
DR P-PSDB; ABG68755.

Novel cytochrome P450 protein in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, useful as medicament, and in evaluating drug metabolism, in drug design and drug screening.

Disclosure; Fig 13; 131pp; English.

XX The invention describes a cytochrome P450 protein (I) in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, as well as sub fragments, variants and multiples of (I) having essentially the same characteristics. (I) is useful as a medicament, and for evaluating drug metabolism, in drug design, and drug screening, and in tests for adjusting the dose of drugs. This sequence encodes a novel cytochrome P450 protein

XX Sequence 1062 BP; 318 A; 239 C; 216 G; 289 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 6; Length 1062;  
Best Local Similarity 50.7%; Pred. No. 0.014;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGCCGATGAGAGGACTCGCTCAGATCAGTAACCTGTTATTTCTGATG 1056  
Db 413 AAGCTCTGCCGATGAGAGGCTCGTGGCCCAATCAATTATCTTTATTTTCTGGCTATG 472  
QY 1057 AAATCTCTTCGACGCTTTCATGATGATGTTTCCCGACTCTCAGAACAAAGCCGTC 1116  
Db 473 AAACACGAGCAGTGTCTCTCTTCATTATGATGAAGTGGCCACTCACCTGATGTC 532  
QY 1117 AGGATAAATCTCGAAGAAATTTGTTCAGATCGACACCGATATGCTACGCTAGACGAC 1176  
Db 533 AGCAGAAATCTGAGGAGAAATTTGATGATGATTTTACCCTAATGAGCACCACCTATG 592

QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTCGCTAGACC 1230  
Db 593 ATACTGTCTACAGATGGAGTATCTTGACATGGTGTGATGAACGCTCAGATTATTC 652  
QY 1231 CTCCTAGTCGATGCTTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
Db 653 CAATTGCTATGAGACTTGAAGGGTCTGCAAAAAGATGTTGATCA 700

RESULT 7  
ABK97693  
ID ABK97693 standard; DNA; 1365 BP.

XX  
AC ABK97693;  
XX  
DT 07-OCT-2002 (first entry)

XX DE DNA encoding cytochrome P450 CYP3A43 exon 1-CYP3A4 exon 4-13.

XX Cytochrome P450; CYP3AP1; CYP3AP2; CYP3A43; CYP3A4; CYP3A5; CYP3A7;  
KW drug metabolism; drug design; drug screening; gene; ds.

XX OS Homo sapiens.

XX PN WO200244213-A1.

XX PD 06-JUN-2002.

XX PF 28-NOV-2001; 2001WO-SE002631.

XX PR 28-NOV-2000; 2000SE-00004366.

PR 11-JUN-2001; 2001SE-00002061.

XX (ZAPH/) ZAPHIROPOULOS P G.

PA (FINT/) FINTA C.

XX Zaphiropoulos PG, Finta C;

XX WPI; 2002-557532/59.

DR P-PSDB; ABG68754.

Novel cytochrome P450 protein in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, useful as medicament, and in evaluating drug metabolism, in drug design and drug screening.

XX Claim 9; Fig 12; 131pp; English.

XX The invention describes a cytochrome P450 protein (I) in which CYP3A43 exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, as well as sub fragments, variants and multiples of (I) having essentially the same characteristics. (I) is useful as a medicament, and for evaluating drug metabolism, in drug design, and drug screening, and in tests for adjusting the dose of drugs. This sequence encodes a novel cytochrome P450 protein

XX Sequence 1365 BP; 406 A; 299 C; 294 G; 366 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 6; Length 1365;  
Best Local Similarity 50.7%; Pred. No. 0.016;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGCCGATGAGAGGACTCGCTCAGATCAGTAACCTGTTATTTCTGATG 1056  
Db 716 AAGCTCTGCCGATGAGAGGCTCGTGGCCCAATCAATTATCTTTATTTTCTGGCTATG 775  
QY 1057 AAATCTCTTCGACGCTTTCATGATGATGTTTCCCGACTCTCAGAACAAAGCCGTC 1116  
Db 776 AAACACGAGCAGTGTCTCTCTTCATTATGATGAAGTGGCCACTCACCTGATGTC 835  
QY 1117 AGGATAAATCTCGAAGAAATTTGTTCAGATCGACACCGATATGCTACGCTAGACGAC 1176  
Db 836 AGCAGAAATCTGAGGAGAAATTTGATGATGATTTTACCCTAATGAGCACCACCTATG 895



QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGAAGAGTCTCTTCGCTAGACC 1230  
 Db 896 ATACTGTGCTACAGATGGAGTATCTTGACATGGTGGTGAATCAAAACGCTCAGATTATCC 955  
 QY 1231 CTCCTAGTCCGATCTAACCCTGAATCTTAAAGGATGAAGACTTCA 1278  
 Db 956 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGAGATCA 1003

RESULT 8  
 ADJ87520  
 ID ADJ87520 standard; DNA; 1458 BP.  
 XX  
 AC ADJ87520;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human cytochrome P450 3A4 coding sequence.  
 XX  
 KW ds; gene; cytochrome P450; crystallography.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.

Key Location/Qualifiers  
 CDS 1..1458  
 FT /tag= a  
 FT /product= "human cytochrome P450 3A4"

WO2003102192-A1.

11-DEC-2003.

30-MAY-2002; 2002WO-GB002668.

30-MAY-2002; 2002WO-GB002668.

(ASTE-) ASTEX TECHNOLOGY LTD.

Cosme J, Ward A, Vuillard L, Williams P, Hamilton B;

WPI; 2004-043119/04.

P-PSDB; ADJ87521.

Purifying a cytochrome P450, for NMR studies and high-throughput screening methods to discover drugs, comprises suspending cells expressing P450 molecule in a salt buffer, lysing the cells and providing a high-salt-detergent lysate.

Example 4; SEQ ID NO 7; 77pp; English.

The invention relates to a method of purifying a cytochrome P450 by expressing in a host cell culture a cytochrome P450 molecule, recovering the cells from the culture and suspending the cells in a salt buffer having a conductivity of 12-110 mS/cm, lysing the cells and removing cell debris to provide a high-salt lysate, adding to the lysate a detergent to provide a high-salt-detergent lysate, and recovering the P450 from the lysate. Methods of preparing and purifying cytochrome P450 proteins are useful for X-ray crystallographic studies and crystallographic screening of small molecules bound to P450, or for NMR studies and high-throughput screening methods to discover drugs or analyze the interaction of drugs with P450 molecules. This sequence corresponds to the coding sequence for the human cytochrome P450 3A4.

Sequence 1458 BP; 436 A; 313 C; 316 G; 393 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 12; Length 1458;

Best Local Similarity 50.7%; Pred. No. 0.016;

Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGGATGAGGAGTACTCGCTCAGATCAAGTAACTGTTATTTGCTGGATATG 1056

Db 797 AAGCTCTGCGATCTGGAGCTCGTGGCCAAATCAATTAATCTTTATTTTGGCTATG 856

RESULT 9  
 ADO33767  
 ID ADO33767 standard; DNA; 1458 BP.  
 XX  
 AC ADO33767;

29-JUL-2004 (first entry)

Human CYP3A4 truncated DNA with synthetic N-terminal and His tag.

cytochrome P450; CYP3A4; protein co-ordinate data; cytostatic; cancer; hydrophobic transmembrane domain truncated; ds; gene; human.

Homo sapiens.

Synthetic.

Key Location/Qualifiers

CDS 1..1458

FT /tag= a

FT /product= "Human cytochrome P450 CYP3A4 truncated protein with synthetic N-terminal and His tag"

WO2004038015-A1.

06-MAY-2004.

24-OCT-2003; 2003WO-GB004598.

25-OCT-2002; 2002US-0421063P.

19-JUN-2003; 2003US-0479448P.

(ASTE-) ASTEX TECHNOLOGY LTD.

Tickle IJ, Vonnrhein C, Williams PA, Jhoti H, Kirtson SB;

WPI; 2004-440452/41.

P-PSDB; ADO33768.

Obtaining a representation of the 3-D structure of cytochrome P450 3A4 crystals, by providing data of structure factors used to generate electron density map of crystal structure and constructing electron density map of obtained data.

Example; SEQ ID NO 1; 357pp; English.

The invention relates to a novel method for obtaining a representation of the 3-dimensional structure of a crystal of cytochrome P450 (CYP)3A4 which involves providing the data of structural factors and phases used to generate an electron density map of 3A4 crystal structures as given in specification and constructing an electron density map of the structural factors and phases. The method of the invention has cytostatic applications and may be useful for obtaining a representation of the 3-dimensional structure of a crystal of CYP3A4, where the crystal structure is useful in modelling the interaction of a compound with the protein and in drug design. Such information may be utilised on order to generate a

CC composition to treat cancer. The current sequence is that of the human  
 CC CYP3A4 N-terminal truncated DNA of the invention which has a synthetic N-  
 CC terminal region in place of the hydrophobic transmembrane domain in order  
 CC to aid *Escherichia coli* expression and solubility, in addition to a C-  
 CC terminal His tag which facilitates purification.

XX SQ Sequence 1458 BP; 436 A; 313 C; 316 G; 393 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 12; Length 1458;  
 Best Local Similarity 50.7%; Pred. No. 0.016;  
 Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATCGAGGAGTACTCGCTCAGATCACTGTTATTTTCTGGATATG 1056

DB 797 AAGCTCTGTCGATCGAGGAGTACTCGCTCAGATCACTGTTATTTTCTGGATATG 856

QY 1057 AAATCTTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGCAAAAGCCGTTTC 1116

DB 857 AAACCCAGAGCAGTGTCTCTCTCCATGATGTAAGTGGCCACTCACCCTGATGTC 916

QY 1117 AGGATAAACTTCGAGAAAGAAATTTGTCAGATCGACCGGATATGCTTACGCTAGACGAAC 1176

DB 917 AGCAGAACTCGAGGAGGAATTCATGCAGTTTTTACCCCAATAGGCCACCCACCTATG 976

QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGCTAGACC 1230

DB 977 ATACTGTCTACAGATGGAGTATCTTGACATGGTGTGAATGAACCGCTCAGATTATTC 1036

QY 1231 CTCCTAGTCCGATGCTTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278

DB 1037 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAAGATGTTGAGATCA 1084

# RESULT 10

ID AAQ87717 standard; cDNA; 1512 BP.

XX AAQ87717;

XX 25-MAR-2003 (revised)

XX 14-NOV-1995 (first entry)

XX Human cytochrome P450 molecular species 3A4 cDNA coding region.

XX Human cytochrome P450; amplification; PCR; primer; expression vector;  
 KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;  
 KW carcinogen; mutagen; liver metabolism; ds.

XX Homo sapiens.

XX EP644267-A2.

XX 22-MAR-1995.

XX 20-JUL-1994; 94EP-00111298.

XX 20-JUL-1993; 93JP-00201120.

XX 21-JUL-1993; 93JP-00180246.

XX 30-JUL-1993; 93JP-00208279.

XX (SUMO ) SUMITOMO CHEM CO LTD.

XX Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;

XX WPI: 1995-116991/16.

XX F-PSDB; AAR72363.

XX Evaluation of safety of a chemical cpd. - using recombinant yeast  
 PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.

XX Example; Page 31-33; 124pp; English.

XX The nucleotide sequence of the cDNA coding region for the human

CC cytochrome P450 species 3A4. The gene encodes a protein of 503 amino  
 CC acids. The cDNA was amplified by PCR using the primers AAQ87743-5. The  
 CC product was cloned into the yeast expression vectors pAAH5N or pAHRH to  
 CC produce the vectors p3A4 for the expression of the cytochrome P450 alone  
 CC or p3A4R for co-expression with the yeast NADPH-P450 reductase. The  
 CC vectors are used in a method for evaluating the safety of a chemical  
 CC compound by reacting the chemical compound with recombinantly produced  
 CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),  
 CC 2E1 (AAQ87716), or 3A4 or their auxiliary species and variants (AAQ87718-  
 CC 32), and yeast NADPH-P450 reductase, either as a fused protein or in cell  
 CC extracts, and analysing the resulting metabolite to assess the safety of  
 CC the chemical compound. The method is useful for determining whether the  
 CC chemical compound, or its metabolite, will be converted into a  
 CC carcinogenic or mutagenic form through metabolism in the liver. (Updated  
 CC on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 2; Length 1512;

Best Local Similarity 50.7%; Pred. No. 0.016;

Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGATCGAGGAGTACTCGCTCAGATCACTGTTATTTCTGGATATG 1056

DB 863 AAGCTCTGTCGATCGAGGAGTACTCGCTCAGATCACTGTTATTTCTGGATATG 922

QY 1057 AAATCTTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGCAAAAGCCGTTTC 1116

DB 923 AAACCCAGAGCAGTGTCTCTCTCCATGATGTAAGTGGCCACTCACCCTGATGTC 982

QY 1117 AGGATAAACTTCGAGAAAGAAATTTGTCAGATCGACCGGATATGCTTACGCTAGACGAAC 1176

DB 983 AGCAGAACTCGAGGAGGAATTCATGCAGTTTTTACCCCAATAGGCCACCCACCTATG 1042

QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGCTAGACC 1230

DB 1043 ATACTGTCTACAGATGGAGTATCTTGACATGGTGTGAATGAACCGCTCAGATTATTC 1102

QY 1231 CTCCTAGTCCGATGCTTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278

DB 1103 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAAGATGTTGAGATCA 1150

# RESULT 11

AAAT17399

ID AAAT17399 standard; cDNA; 1512 BP.

XX AAAT17399;

XX 01-AUG-1996 (first entry)

XX Human derived cytochrome P4503A4 cDNA.

XX Human derived cytochrome; P4503A4; commercial cDNA library; yeast;  
 KW transfection; recombinant production; expression vector; mammal;  
 KW immunisation; sensitisation; antibody; determination; detection;  
 KW non-cross reactive; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 1..1512

XX /\*tag= a

XX JP08027197-A.

XX 30-JAN-1996.

XX 13-JUL-1994; 94JP-00161552.

XX 13-JUL-1994; 94JP-00161552.

XX (SUMO ) SUMITOMO CHEM CO LTD.

```

XX WPI; 1996-136338/14.
DR P-PSDB; AAR81464.
XX
XX Antibody recognising human derived cytochrome P450A4 - allows specific
PT detection of cytochrome P450 species in humans.
XX
XX Example 1; Page 10-12; 13pp; Japanese.
XX
XX The present sequence encodes the human derived cytochrome (HDC) P450A4,
CC which was obtd. from a commercial cDNA library. Yeast were transfected
CC with an expression vector congt. the HDC cDNA, cultured and then
CC disrupted to give a microsomal fraction. The HDC was purified from the
CC fraction, and used to immunise and sensitise a mammal. Blood was drawn
CC from the mammal, and an anti-HDC antibody isolated. The antibody obtd.
CC recognises HDC P450A4, partic. at a serum dilution rate of 1:10000, and
CC is substantially without cross reaction to other HDC P450 spp
XX
XX Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 U; 0 Other;
SQ
Query Match 2.5%; Score 48.4; DB 2; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.016;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
QY 997 AAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACTGTTATTTGCTGGATATG 1056
Db 863 AAGCTCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACTGTTATTTGCTGGATATG 922
QY 1057 AAACCTTTCTCGACAGTCTTGACATGGATGTTTCACCGACTCTCAGAACAAAGCGTTC 1116
Db 923 AAACACGAGCAGTGTCTCTCTTCATTTATGATGAACCTGGCCACTCACCTGATGTCC 982
QY 1117 AGGATAAATTCGAGAGAAATTTGTTCAGATCGACAGGATATGCTTACCTAGACGAAC 1176
Db 983 AGCAGAACTGCAGAGGAAATTTGATCGAGTTTACCCATTAAGGCACACCCACTATG 1042
QY 1177 TTAATGCGTTC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTGCTAGAAC 1230
Db 1043 ATACTGTGCTACAGATGAGATCTCTGACATGCTGTGTAATGAACGCTCAGATTATTC 1102
QY 1231 CTCCTAGTCCGTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
Db 1103 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGAGATCA 1150
RESULT 12
AAT28383
ID AAT28383 standard; DNA; 1512 BP.
XX
XX AAT28383;
AC
XX 11-OCT-1996 (first entry)
DT
XX Human cytochrome P450 molecular species 3A4 gene.
DE
XX
XX Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
XX carcinogenic; ds.
XX
XX Homo sapiens.
OS
XX
XX JF08056695-A.
PN
XX
XX 05-MAR-1996.
PD
XX 15-JUL-1994; 94JP-00164184.
XX
XX 20-JUL-1993; 93JP-00201120.
PR
XX 30-JUL-1993; 93JP-00208279.
PR
XX 17-JUN-1994; 94JP-00136053.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
PA

```

```

XX WPI; 1996-182311/19.
DR P-PSDB; AAR93170.
XX
XX Novel method for the evaluation of the safety of a cpd. - using a human
PT cytochrome P450 and yeast NADPH reductase to determine whether the
PT analyte cpd. is detoxified or metabolised to a carcinogen.
XX
XX Example 1; Page 24-26; 7app; Japanese.
XX
XX This is the nucleotide sequence of the human cytochrome P450 molecular
CC species 3A4 gene which encodes a protein of 503 amino acids. The gene was
CC amplified from a human liver derived cDNA library as 2 fragments of 0.6
CC and 0.9 kb using primers AAT26933-6. The prod. was cloned into the yeast
CC expression vector pAAH5N to generate plasmid p3A4 for prodn. of the
CC cytochrome only or into the vector pAHR to generate the plasmid p3A4R
CC for co-prodn. with the yeast NADPH-P450 reductase. The sequence is placed
CC under control of the yeast ADH gene promoter and terminator. The vectors
CC are used in a method for evaluating the safety of a cpd. by reacting the
CC test cpd. with recombinantly produced human cytochrome P450 mol. species
CC 1A2 (AAT28380), 2C9 (AAT28381), 2E1 (AAT28382), 3A4 or their variants
CC (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC protein or as a cell extract) and analysing the resultant metabolite. The
CC cpd. is considered "safe" if it is detoxified or not rendered
CC carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC carcinogenic cpd
XX
XX Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 U; 0 Other;
SQ
Query Match 2.5%; Score 48.4; DB 2; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.016;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
QY 997 AAAGCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACTGTTATTTGCTGGATATG 1056
Db 863 AAGCTCTGTCGATGAGAGGTACTCGCTCAGATCAGTAACTGTTATTTGCTGGATATG 922
QY 1057 AAACCTTTCTCGACAGTCTTGACATGGATGTTTCACCGACTCTCAGAACAAAGCGTTC 1116
Db 923 AAACACGAGCAGTGTCTCTCTTCATTTATGATGAACCTGGCCACTCACCTGATGTCC 982
QY 1117 AGGATAAATTCGAGAGAAATTTGTTCAGATCGACAGGATATGCTTACCTAGACGAAC 1176
Db 983 AGCAGAACTGCAGAGGAAATTTGATCGAGTTTACCCATTAAGGCACACCCACTATG 1042
QY 1177 TTAATGCGTTC-----CTTATCTCGAAGCGTTTGTAAAGAGTCTCTTTGCTAGAAC 1230
Db 1043 ATACTGTGCTACAGATGAGATCTCTGACATGCTGTGTAATGAACGCTCAGATTATTC 1102
QY 1231 CTCCTAGTCCGTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
Db 1103 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGAGATCA 1150
RESULT 13
ABK97692
ID ABK97692 standard; DNA; 1512 BP.
XX
XX ABK97692;
AC
XX 07-OCT-2002 (first entry)
DT
XX
XX DNA encoding cytochrome P450 CYP3A43 exon 1-CYP3A4 exon 2-13.
DE
XX
XX Cytochrome P450; CYP3AP1; CYP3AP2; CYP3A43; CYP3A4; CYP3A5; CYP3A7;
KW drug metabolism; drug design; drug screening; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200244213-A1.
PN
XX
XX 06-JUN-2002.
PD
XX
XX

```

XX PF 28-NOV-2001; 2001WO-SE002631.  
XX XX  
XX PR 28-NOV-2000; 2000SE-00004366.  
XX PR 11-JUN-2001; 2001SE-00002061.  
XX XX  
XX PA (ZAPH/) ZAPHIROPOULOS P G.  
XX PA (FINT/) FINTA C.  
XX XX  
XX PI Zaphirooulos PG, Finta C;  
XX DR WPI; 2002-557532/59.  
XX DR P-PSDB; ABG68753.  
XX XX  
XX PT Novel cytochrome P450 protein in which CYP3A43 exon 1 is joined to sets  
XX PT of CYP3A4 or CYP3A5 exons, useful as medicament, and in evaluating drug  
XX PT metabolism, in drug design and drug screening.  
XX XX  
XX PS Claim 9; Fig 11; 131pp; English.

CC The invention describes a cytochrome P450 protein (I) in which CYP3A43  
CC exon 1 is joined to sets of CYP3A4 or CYP3A5 exons, as well as sub  
CC fragments, variants and multiples of (I) having essentially the same  
CC characteristics. (II) is useful as a medicament, and for evaluating drug  
CC metabolism, in drug design, and drug screening, and in tests for  
CC adjusting the dose of drugs. This sequence encodes a novel cytochrome  
CC P450 protein

XX Sequence 1512 BP; 447 A; 324 C; 328 G; 413 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 6; Length 1512;  
Best Local Similarity 50.7%; Pred. No. 0.016;  
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTCTCGATGAGGAGTACTCGCTCAGATCAGTAACCTGTATTTCCTGGATATG 1056  
Db 863 AAGCTCTCTCGATCTGGAGCTCGTGCCCAATCAATTATCTTTATTTTGTGGCTATG 922

QY 1057 AAACCTCTTCGACAGTCTTGACATGGATGTTTACCAGACTCTCAGAAGACAAGCCGTTTC 1116  
Db 923 AAACACAGAGAGTGTCTCTCTTCAITATGATGAACCTGGCACTCACCTCATGTGCC 982

QY 1117 AGGATAAACTTCGAGAAGAAAATTTGTGCAGATCGACACGGATATGCCTACGCTAGACGAC 1176  
Db 983 ACCAGAAACTGCAGGAGGAATTTGATGCAGTTTTTACCCTAATAGGCACCAACCCCTATG 1042

QY 1177 TTAATGCGTTGC-----CTTATCTCGAACGTTTGTTAAGGAGTCTCTTCGCTAGACC 1230  
Db 1043 ATACTGTCTACAGATGGAGTATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 1102

QY 1231 CTCCTAGTCCGATGCTAACTGCTTAAAGGATGAAGACTTCA 1278  
Db 1103 CAATTGCTATGAGACTTGAGGGGTCTGCAAAAAGATGTTGAGATCA 1150

RESULT 14  
ID ACF06054  
XX ACF06054 standard; cDNA; 1512 BP.  
XX AC ACP06054;  
XX OS Homo sapiens.  
XX DT 22-SEP-2003 (first entry)  
XX DE Human cytochrome P450 3A4 encoding cdna.  
XX KW Human; protein array; protein moiety; phenotype; drug discovery;  
XX KW naturally occurring variant; pharmacogenomic; diagnostic;  
XX KW parallel analysis; tumour suppressor; p53; cytochrome P450; gene; ss.  
XX OS Homo sapiens.  
XX PN WO2003048768-A2.  
XX IN 11 JUN 2003

KW/ protein array; protein moiety; drug metabolising enzyme; DME;  
 KW drug metabolism; drug toxicity; cytotoxicity; drug metabolite;  
 KW metabolic pathway; human; cytochrome; enzyme; P450; gene; ss.  
 XX Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 1..1512  
 FT /\*tag= a  
 FT /product= "cytochrome P450 enzyme 3A4"

XX WO2004025244-A2.

XX PD 25-MAR-2004.

XX PF 16-SEP-2003; 2003WO-IB005258.

XX PR 16-SEP-2002; 2002US-0410815P.

XX PR 05-DEC-2002; 2002US-00313963.

XX PR 05-DEC-2002; 2002WO-GB005499.

XX PA (SENS-) SENSE PROTEOMIC LTD.

XX PI Boutell JM, Godber BLJ, Hart DJ, Bockett NA, Kozlowski R;

XX DR WPI; 2004-270121/25.

XX DR P-PSDB; ADL18672.

XX New protein array comprising a surface having spatially defined locations  
 PT containing drug metabolizing enzymes, examining gender and ethnicity-  
 PT related differences in drug metabolism or cytotoxicity of drug  
 PT metabolites.

XX Example 3; Fig 3A; 72pp; English.

XX The present invention describes a protein array comprising a surface  
 CC having spatially defined locations where at each location there are  
 CC deposited at least two protein moieties capable of forming a complex,  
 CC where the complex is transiently formed and where the protein moieties  
 CC act sequentially on a substrate of interest and are derived from one or  
 CC more drug metabolising enzymes (DMEs). Also described: (1) a method of  
 CC making a protein array; (2) an array made by the method of (1); (3) a  
 CC method of screening a set of protein moieties for molecules which  
 CC interact with one or more proteins; (4) a method of simultaneously  
 CC determining the relative properties of members of a set of protein  
 CC moieties; and (5) a method of expressing and purifying a DME. The protein  
 CC array is useful in examining gender differences in drug metabolism,  
 CC ethnicity-related differences in drug metabolism and toxicity between two  
 CC or more mammalian species, e.g. human and rat and cytotoxicity of drug  
 CC metabolites, in defining and quantifying metabolic pathways for small  
 CC molecules, in screening of compounds that binds and inhibits individual  
 CC DMEs and in analysing induction of P450 expression by one or more  
 CC compounds of interest and the effects of mutation on the activity of a  
 CC DME of interest. The present sequence encodes human cytochrome P450  
 CC enzyme 3A4, which is used in the exemplification of the present  
 CC invention.

XX SQ Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 U; 0 Other;

Query Match 2.5%; Score 48.4; DB 12; Length 1512;  
 Best Local Similarity 50.7%; Pred. No. 0.016;  
 Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTGTCGAGTGGAGGAGTCTCGTCAGATCAGTAACTGTTATTGCTGGATATG 1056

Db 863 AAGCTCTGTCGATCTGGAGCTGTCGGCCCAATCAATATCTTTATTTTGGCTATG 922

QY 1057 AAACCTCTTCGACAGCTTTCGATGTTTCACCGACTCTCAGAGACAAACCGCTTC 1116

Db 923 AAACCCAGGACAGTGTCTCTCCTTCATTATGATGAATGCGCACTCACCCCTGATGTC 982

QY 1117 AGGATAACTTCGAGAGAGATTTGTCAGATCGACCGGATATGCTACGCTAGACGAC 1176

Db 983 AGCAGAACTGCAGGAGGAAATTGATGCACTTTTACCCCAATAAGGCACCACCTATG 1042  
 QY 1177 TTAATCCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTGTCTAGACC 1230  
 Db 1043 ATACTGTGCTACAGATGGAGTATCTTGACATGCTGCTGAATGAAAGCTTCAGATTATTC 1102  
 QY 1231 CTCCTAGTCCGTATGCTAAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278  
 Db 1103 CAATTGCTATGAGACTTGGAGGGTCTGCAGAAAAGAGATGTTGAGATCA 1150

Search completed: December 14, 2004, 02:38:08  
 Job time : 898.808 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2004, 22:14:22 / Search time 8080.28 Seconds  
(without alignments)  
11306.990 Million cell updates/sec

Title: US-10-066-007A-2

Perfect score: 1932

Sequence: 1 gaattcgacagagccacc.....agccggctcgtgcgaattc 1932

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_bt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1932	100.0	1932	6	AR203449 Sequence
2	1932	100.0	1932	6	AX034665 Sequence
3	353.2	18.3	3969	6	AR203450 Sequence
4	353.2	18.3	3969	6	AX034666 Sequence
5	56.8	2.9	2699	5	BC077479 Xenopus l
6	55	2.8	2514	5	EC060436
7	53.4	2.8	1285	5	BC060496 Xenopus l
8	51.8	2.7	2000	6	BX931257 Gallus ga
9	48.4	2.5	1062	9	AJ563377 Homo sapi
10	48.4	2.5	1197	6	CQ717693 Sequence
11	48.4	2.5	1365	9	HSAS63376
12	48.4	2.5	1458	6	CQ812569
13	48.4	2.5	1458	6	AX960331 Sequence
14	48.4	2.5	1512	6	CQ846981
15	48.4	2.5	1512	6	E10636
16	48.4	2.5	1512	6	E10855
17	48.4	2.5	1512	6	AR399323
18	48.4	2.5	1512	6	AY582531
19	48.4	2.5	1512	9	HSAS63375

20	48.4	2.5	1807	9	BC069418
21	48.4	2.5	2011	6	AX408939
22	48.4	2.5	2011	9	HUMCYPLP
23	48.4	2.5	2011	9	HUMP450M
24	48.4	2.5	2032	6	CQ717692 Sequence
25	48.4	2.5	2036	9	HUMCYPNO
26	48.4	2.5	2059	6	AR380517
27	48.4	2.5	2059	9	HUMCYPNOA
28	48.4	2.5	2059	11	G18329
29	48.4	2.5	2759	6	AR222892
30	48.4	2.5	2759	9	HSRCYP3
31	48.4	2.5	2849	6	AX409463
32	48.4	2.5	2849	9	HUMCYTIIIA
33	48	2.5	154452	2	AC137912
34	48	2.5	172285	2	AC147670
35	48	2.5	211728	2	BS000602
36	47.4	2.5	1410	8	AB084385
37	47.4	2.5	1684	10	BC040779
38	47.4	2.5	1871	8	AK072295
39	47.4	2.5	1900	8	AK101670
40	47.4	2.5	2160	6	AR409280
41	47.4	2.5	2160	6	BD140525
42	47.2	2.4	3588	9	HSB802481
43	47.2	2.4	172285	2	AC147670
44	47	2.4	445	6	CQ507205
45	46.8	2.4	1512	9	AY334551

## ALIGNMENTS

RESULT 1  
AR203449  
LOCUS AR203449 1932 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 2 from patent US 6365386.  
ACCESSION AR203449  
VERSION AR203449.1 GI:21499840  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1932)  
AUTHORS Hoshino, T., Ojima, K. and Setoguchi, Y.  
TITLE Astaxanthin synthase  
JOURNAL Patent: US 6365386-A 2 02-APR-2002;  
FEATURES Location/Qualifiers  
1..1932  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Query Match 100.0%; Score 1932; DB 6; Length 1932;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGAGCCACCTTCTCCATATGTTTCATCTTGGTCTTGTCTCAGAGTG 60  
DB 1 GAATTCGGCAGAGCCACCTTCTCCATATGTTTCATCTTGGTCTTGTCTCAGAGTG 60  
QY 61 CTTTAGGCTGGTCTTCTTCATGGGCATCCATAGCGTCTTTCAGTCTTTACCTCGCTC 120  
DB 61 CTTTAGGCTGGTCTTCTTCATGGGCATCCATAGCGTCTTTCAGTCTTTACCTCGCTC 120  
QY 121 CGAGCGATCTTCTACTGTATTAACCTTCAGGCCCGGAATCATACCACTTTACAGGCA 180  
DB 121 CGAGCGATCTTCTACTGTATTAACCTTCAGGCCCGGAATCATACCACTTTACAGGCA 180  
QY 181 ATTTTITAGACATCTTCTCAGCTCGTAGAGAGCATGCGAAGTACAGAGAAAAT 240  
DB 181 ATTTTITAGACATCTTCTCAGCTCGTAGAGAGCATGCGAAGTACAGAGAAAAT 240  
QY 241 ACGGAGACCTCGGTTCTGGATCGGTGAGCACCGCTTGAATTCGACCGATC 300





DKMDEARAKDMVAGSAGKATRELETESGVVDKDVGRATLDYMALAGFDYKSDSLQ  
 NKTNELYVAFVGLDGFAPTLDSFKAINMDFVPIFRMKRHEIPLTCOGAVSRVGI  
 SLMEQKQAVGSASDQAVDKKDVQGRDILSLVRANIANLPESQKLSDEVLQAIS  
 NLLFAGYETSSTVLTMFHRLESDEKAVQDKLREBIQCIDTDMPTLDELNALPYLEAFV  
 KESLRLDPPSPYANRECLKDEDFIPLAEPVTRGDSGVINEVRITFGWMLPLFNINR  
 SKFIYGEDABEFRRPEMLDVTDSLNGTEAPYHOASFIISGPRACFQWRPAVAEMKAF  
 LFWTLRRVQPEPIIISHPEYEHITLISREPRIVGREKGYQMRLOKQVPE"  
 1871

polya\_site  
 ORIGIN

Query Match 100.0%; Score 1932; DB 6; Length 1932;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAATTCGCGACGAGCCACCTACTTCTCCATATGTCATCTTGGTCTTGTCTCACAGTG	60
DB	1	GAATTCGCGACGAGCCACCTACTTCTCCATATGTCATCTTGGTCTTGTCTCACAGTG	60
QY	61	CTTTAGGCTGGCTGCTTCTCATGGCATCCATAGCGTCTTTCAGTCTTTTACCTCGCTC	120
DB	61	CTTTAGGCTGGCTGCTTCTCATGGCATCCATAGCGTCTTTCAGTCTTTACCTCGCTC	120
QY	121	CGAGCGATCTCACTGATTAACCTTCAGGCCCGGAATCATACCACTACTTTACAGGCA	180
DB	121	CGAGCGATCTCTCACTGATTAACCTTCAGGCCCGGAATCATACCACTACTTTACAGGCA	180
QY	181	ATTTTTAGACATCTCTCAGCTCGGTACAGGTGAAGCATCGGAAGTACAGAGAAAAT	240
DB	181	ATTTTTAGACATCTCTCAGCTCGGTACAGGTGAAGCATCGGAAGTACAGAGAAAAT	240
QY	241	ACGGAAGCACCTCCGGTTTCTGGGATCGCTGGAGCACCGCTTTTGAATCTGACCGATC	300
DB	241	ACGGAAGCACCTCCGGTTTCTGGGATCGCTGGAGCACCGCTTTTGAATCTGACCGATC	300
QY	301	CGAAAGTCTTCAACCATGTGATGAAGAAGCTCAGCATATCCGAAACCTGGTATGCCCG	360
DB	301	CGAAAGTCTTCAACCATGTGATGAAGAAGCTCAGCATATCCGAAACCTGGTATGCCCG	360
QY	361	CTCAGTGTCTCAGAAATGCTACCGAGATGGTGTGTTAGCGCGGAAGGTGAAGTCTATA	420
DB	361	CTCAGTGTCTCAGAAATGCTACCGAGATGGTGTGTTAGCGCGGAAGGTGAAGTCTATA	420
QY	421	AGCGACATCGAAGATCATGATCCCTCTCTGTCGGCTCAGGCCGTTAAGTCGATGGTCC	480
DB	421	AGCGACATCGAAGATCATGATCCCTCTCTGTCGGCTCAGGCCGTTAAGTCGATGGTCC	480
QY	481	CAATTTTCTTAGAAAAAGGTATGAACTTTGTCGACAAAGATGATGGAGATGGCGCTGAGA	540
DB	481	CAATTTTCTTAGAAAAAGGTATGAACTTTGTCGACAAAGATGATGGAGATGGCGCTGAGA	540
QY	541	AGGATATGCCGCTGGGAGTCGCGGTGAAAAGAGGCAACAGACTCGAGACCGAAG	600
DB	541	AGGATATGCCGCTGGGAGTCGCGGTGAAAAGAGGCAACAGACTCGAGACCGAAG	600
QY	601	GAGTCGATGTAAGGATGGGTCCGTCCAGTCTCTGGACGTCATGGCTCTTTCGAGGAT	660
DB	601	GAGTCGATGTAAGGATGGGTCCGTCCAGTCTCTGGACGTCATGGCTCTTTCGAGGAT	660
QY	661	TTGACTATAAGAGCGACTCGCTCCAGAACCAAGCAATGAGCTCTATGTCGCTTTTGTGCG	720
DB	661	TTGACTATAAGAGCGACTCGCTCCAGAACCAAGCAATGAGCTCTATGTCGCTTTTGTGCG	720
QY	721	GACTTACCGATGGGTTTGTCTTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTG	780
DB	721	GACTTACCGATGGGTTTGTCTTACCTTGGACTCGTTCAAGGCTATCATGTGGGATTTTG	780
QY	781	TACCTTACTTCCGAATATGAACCGGAGCATGAGATACCTTTGACTCAAGGATTTAGCAG	840
DB	781	TACCTTACTTCCGAATATGAACCGGAGCATGAGATACCTTTGACTCAAGGATTTAGCAG	840
QY	841	TTTCCCGACGATTTGGGATCGAGCTTATGAGCAAAAGAGCGCGCTGCTGGCTCAG	900
DB	841	TTTCCCGACGATTTGGGATCGAGCTTATGAGCAAAAGAGCGCGCTGCTGGCTCAG	900

QY	901	CTTCCGATCAGGCTGTGTGATAAAAAAGGATGTTCAAGGTCGGGATATCCTAAGTCTCCTAG	960
DB	901	CTTCCGATCAGGCTGTGTGATAAAAAAGGATGTTCAAGGTCGGGATATCCTAAGTCTCCTAG	960
QY	961	TGAGAGCAACATCGCGCCGCAACCTGCTGCAATCTCAAAAGCTGTCGGATGAGGAGTAC	1020
DB	961	TGAGAGCAACATCGCGCCGCAACCTGCTGCAATCTCAAAAGCTGTCGGATGAGGAGTAC	1020
QY	1021	TCGCTCAGATCAGTAACTGTTTATTTGCTGGGATATGAAACTTCTTCGACAGTCTTGACAT	1080
DB	1021	TCGCTCAGATCAGTAACTGTTTATTTGCTGGGATATGAAACTTCTTCGACAGTCTTGACAT	1080
QY	1081	GGATGTTTCAACGACTCTCAGAGACAAAGCCGTTTCAAGATAACTTCGAGAGAAATTT	1140
DB	1081	GGATGTTTCAACGACTCTCAGAGACAAAGCCGTTTCAAGATAACTTCGAGAGAAATTT	1140
QY	1141	GTGAGATCGACACGGATATGCTTACGCTAGAGCAACTTAATGCTTGCCTTATCTCGAAG	1200
DB	1141	GTGAGATCGACACGGATATGCTTACGCTAGAGCAACTTAATGCTTGCCTTATCTCGAAG	1200
QY	1201	CGTTTGTAAAGAGTCTTCTGCTTAGACCTTCTAGTCCGTATGCTAAACCGTGAATGCT	1260
DB	1201	CGTTTGTAAAGAGTCTTCTGCTTAGACCTTCTAGTCCGTATGCTAAACCGTGAATGCT	1260
QY	1261	TAAAGGATGAAGACTTCAATCCCACTTCCGAGCCCTGCTATTCGAGATGGTGGTCA	1320
DB	1261	TAAAGGATGAAGACTTCAATCCCACTTCCGAGCCCTGCTATTCGAGATGGTGGTCA	1320
QY	1321	TCAAGGATCGCGATCAACGAAAGCAATGCTTCCGTTGTTCAACATCAATC	1380
DB	1321	TCAAGGATCGCGATCAACGAAAGCAATGCTTCCGTTGTTCAACATCAATC	1380
QY	1381	GTTCAAAGTTCATTTATGAGAGATGAGAGAAATTCAGACCGGAGGTTGGTGGAG	1440
DB	1381	GTTCAAAGTTCATTTATGAGAGATGAGAGAAATTCAGACCGGAGGTTGGTGGAG	1440
QY	1441	ACGTAACAGACTCGCTCAACAGTATGAAGCACCCCTATGGAACACGAGCGAGCTTTATCT	1500
DB	1441	ACGTAACAGACTCGCTCAACAGTATGAAGCACCCCTATGGAACACGAGCGAGCTTTATCT	1500
QY	1501	CTGGACCCAGAGCTTGTCTTGGTGGGATTTGCTGTCGCGAGATGAAGGCTTCTTGT	1560
DB	1501	CTGGACCCAGAGCTTGTCTTGGTGGGATTTGCTGTCGCGAGATGAAGGCTTCTTGT	1560
QY	1561	TTGTCACTCTCGGTCCAGTTCGAGCCCATCATCTCTCATCCAGATACGAGCACA	1620
DB	1561	TTGTCACTCTCGGTCCAGTTCGAGCCCATCATCTCTCATCCAGATACGAGCACA	1620
QY	1621	TACCTTGATCATTTTCCGCTCCGATCGTTGTTAGAGAGAGGAGGTTACAGATGC	1680
DB	1621	TACCTTGATCATTTTCCGCTCCGATCGTTGTTAGAGAGAGGAGGTTACAGATGC	1680
QY	1681	GTTTCAGCTCAAGCCGTCGAATGAGTTGATTTCTCATATGTTAAGAGAGTTCTATAT	1740
DB	1681	GTTTCAGCTCAAGCCGTCGAATGAGTTGATTTCTCATATGTTAAGAGAGTTCTATAT	1740
QY	1741	CTGAGATGTGTGACTAGGACAAAGCTTCTTTGTTATCGATTTGTTCTCATACCCGGGC	1800
DB	1741	CTGAGATGTGTGACTAGGACAAAGCTTCTTTGTTATCGATTTGTTCTCATACCCGGGC	1800
QY	1801	AGGCGCTATGACTTCTAGCTGCTTATCGTCCCTTGAGCTCTCTTCTTACCTATATAT	1860
DB	1801	AGGCGCTATGACTTCTAGCTGCTTATCGTCCCTTGAGCTCTCTTCTTACCTATATAT	1860
QY	1861	TATTCATCCGAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT	1920
DB	1861	TATTCATCCGAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT	1920
QY	1921	CGTCCGGAATTC	1932
DB	1921	CGTCCGGAATTC	1932



```

ACCESSION BC077479
VERSION BC077479.1 GI:50603956
KEYWORDS MGC.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 2699)
AUTHORS Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
and Richardson, P.
Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative
Dev. Dyn. 225 (4), 384-391 (2002)
JOURNAL 2 (bases 1 to 2699)
PUBMED Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
AUTHORS Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schaefer, C.F., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL 12477932
PUBMED 3 (bases 1 to 2699)
AUTHORS Klein, S. and Strausberg, R.
REFERENCE Direct Submission
TITLE Submitted (19-JUL-2004) National Institutes of Health, Xenopus Gene
Collection (XGC), National Institute of Child Health and Human
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
20892-7510, USA
REMARK NH-MGC Project
COMMENT Contact: XGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smalish, Jeff Scott, Miranda Tsai, George Yang, Jacquie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 151 Row: f Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1. .2699
/organism="Xenopus laevis"
/mol_type="mRNA"
FEATURES
source
/db_xref="taxon:8355"
/clone="MGC:82514 IMAGE:4963532"
/tissue_type="Spleen, adult Xenopus"
/clone_lib="NICHG_XGC_Spl"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
12. .1502
/codon_start=1
/product="Unknown (protein for MGC:82514)"
/protein_id="AAH77479.1"
/db_xref="GI:50603957"
/translation="MAGSLMKVLSQSSWLPVNVVAIQGFALLCVLLVLLKACALYSRG
RKFTAAALTFPPGPPSHWLYGHVQNRDQKIDRLMVAKKYPNAPFLPIGKFFOTLI
ITHEVAKLVFGRPDKSTGYSLIPWIGKGLVLSGDTWFOHRLITPGPHYDVLK
PVGLISDSTKIMLDKAPSKDEPVELFQHSXMTLDSINKCAFYSQSCQTDIDN
DIQAVYLSWLTQQRITFTFYHNSLIITLSPHGFRRKACRIAHILHTDKVIGKQKL
LSKBEKLVQKKRHLDLFDILLCAKDENGQRSLDEDAEVDITFMFEGHDTSSGIS
WLCMAHTEHQKQCBQBEINEVIGERQTMEDDLNRIPTTMCIKESLRILYPPVSV
SRELAKPTITFDGKSLPAGILIFDLRFSPNSKRSKSHAFVFAAGPRNGTQNFAM
NEKMAVALTLKRFELLPLDLPKPLQLVLSKNGIHVYLKAS"
ORIGIN
Query Match 2.9%; Score 56.8; DB 5; Length 2699;
Best Local Similarity 53.8%; Pred. No. 3.7e-05;
Matches 143; Conservative 0; Mismatches 117; Indels 6; Gaps 1;
QY 996 CAAAGAGCTGCCGATGAGGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTCGATAT 1055
Db 930 CAGAGGCTTCTGATGAGGATATATCGCGTGGAGTGGACACATTCATGTTTGGAGGACAT 989
QY 1056 GAAACTTCTCGACAGTCTTGACATGGATGTTTCCACGACTCTCAGAGACAAAGCCGTT 1115
Db 990 GATACCTACATCTAGTGAATATCTCGATGCTTACATCTGATGCTTACATCAGAGGAC 1049
QY 1116 CAGGATAAATTCGAGAGAGAAATTTGTCAGATCGACAGGATATCGCTACCGTAGAC 1172
Db 1050 CAGCAGAAATGCCAGAGAGATCAATGAAGCTTTGGGGGAGCGCAACCAATGAATGG 1109
QY 1173 ---GAACTTAATGCGTTCCTTATCTCGAAGCGTTTGTAAAGGAGTCTCTTCGTCTAGAC 1229
Db 1110 GATGACCTTAATAGGATACCTTATACCAATGTTGTTAAGGAAAGTCTGCGCTATAC 1169
QY 1230 CTTCTAGTCCGTATGCTAACCGTGA 1255
Db 1170 CCTCCAGTCCCATCAGTATCCAGGGA 1195
RESULT 6
LOCUS BC060496
DEFINITION Xenopus laevis hypothetical protein MGC8821, mRNA (cDNA clone
MGC:168821 IMAGE:4633524), complete cds.
ACCESSION BC060496
VERSION BC060496.1 GI:38051862
KEYWORDS MGC.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 2514)
AUTHORS Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
and Richardson, P.
Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative
Dev. Dyn. 225 (4), 384-391 (2002)
JOURNAL 2 (bases 1 to 2514)
PUBMED Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
AUTHORS Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

```

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaez, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyik, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
3 (bases 1 to 2514)  
Klein, S. and Strausberg, R.  
Direct Submission  
Submitted (27-OCT-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA  
NIH-MGC Project  
Contact: XGC help desk  
Email: cgabps-remail.nih.gov  
Tissue Procurement: Dr. Igor Dawid  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NIISC), Gaithersburg, Maryland  
Web site: <http://www.nisc.nih.gov/>  
Contact: nisc.mgc@nhgri.nih.gov  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.D., Masiello, C., Maskeri, B., Mastrian, S.D., McLooney, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgone, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
Clon distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 128 Row: P Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.  
Location/Qualifiers  
1..2514  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="MGC:68821 IMAGE:4633524"  
/tissue\_type="Lung, adult Xenopus"  
/clone\_lib="NICHD XGC\_Lul"  
/lab\_host="DH10B"  
/notc="Vector: PCMV-SPORT6"  
1..2514  
/gene="MGC68821"  
/db\_xref="LocusID:398964"  
34..1548  
/gene="MGC68821"  
/codon\_start=1  
/product="MGC68821 protein"  
/protein\_id="AAH60496.1"  
/db\_xref="GI:38051863"  
/db\_xref="LocusID:398964"  
/translation="MNLIPLHSTGTWILLAAFLILLYGWIPYGLPKKMGIPGPTPLPIGTFLEBRGLGVQDFTECFKTKTWGLYDGRQVLAINMDPAIKIILVKKCYTNF TNRNFGLSGPLESAITAAEDQWKIRSVLSFTTSGLKEMFQIMKDYSDILVKNL

ODNVKDEPCTMDVIGAYSMVDVITSFVSNIDSLKNPSHPFVHKNLLKGLLNP  
LILVVPFELRIFEGNINFPKQSTQFTNVAVSFRKRRKGDHGTGRVDFLQMW  
DSSVTEGDSNGKALTDVEIMQAQSIIFIPAGETISALSTKFLVCLATHPQVQGL  
HEEDSLFDPKASPTDILQMVEYLDNVIQETLRLPFPFAGLRSVSKQNIENGVSIF  
KGTVTMIPAYLORDPDYWPPEFRFESKENRETHFTFTFLPFGDGRNCIGMRP  
ALLSKVAIVSMQNLQNSVRPCAETLIPMEYSTICVLQPKPKPIVLKFLSRTSAQ"  
148..1521  
/gene="MGC68821"  
/note="P450; Region: Cytochrome P450. Cytochrome P450s are involved in the oxidative degradation of various compounds. Particularly well known for their role in the degradation of environmental toxins and mutagens. Structure is mostly alpha, and binds a heme cofactor"  
/db\_xref="CDD:pfam00067"

## misc\_feature

## ORIGIN

Query Match 2.8%; Score 55; DB 5; Length 2514;  
Best Local Similarity 54.7%; Pred. No. 0.00013;  
Matches 134; Conservative 0; Mismatches 105; Indels 6; Gaps 1;  
QY 997 AAAGCTCTCCGATCAGAGAGTACTCGCTCAGATCAGTAACCTGTTATTTGCTGGATATG 1056  
DB 905 AAGCTCTCAGCAGGTAGAAATTAATGCTCAGTCATATATTTATTTGCTGGATATG 964  
QY 1057 AAACCTCTTCGACAGTCTTGACATGGATGTTTCCCGACTCTTCAGAGACAAACCCGTTCC 1116  
DB 965 AAACAACAAGTACAGCACTGCTTATCTGTTTATTTGTTGCCCCACCCCTGATGTC 1024  
QY 1117 AGGATAAATCTCGAGAGAAATTT-----TGTGAGTCGACACGATGCTTACGCTAG 1170  
DB 1025 AGCAGAGATTCATGAGGAGATAGATTCAATTTCCCTGACAAAGCCAGTCTTACTATG 1084  
QY 1171 ACGAACTTAATCGTTGCTTCTATCTCGAAGCGTTTGTGTAAGAGTCTTCTGCTAGACC 1230  
DB 1085 ACATTTCTGATCAGATGGAATATCTTGATATGGTCATTCAGGAGACCCCTCAGTTGATC 1144  
QY 1231 CTCCT 1235  
DB 1145 CCCCC 1149

## RESULT 7

LOCUS BX931257 1295 bp mRNA linear VRT 02-FEB-2004  
DEFINITION Gallus gallus finished cDNA, clone CHEST68565.  
ACCESSION BX931257  
VERSION BX931257.1 GI:41631785

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## 1..1295

## Location/Qualifiers

## This sequence is from the

## BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection.

## from a library constructed by Elizabeth Bosch. cDNA was prepared

## from RNA extracted from ovary, normalised, and poly A-trimmed.

## EcoRI-NotI cut cDNA was then ligated into the vector. Vector:

## pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia

## coli DH10B.

## colli DH10B.

## Location/Qualifiers

## 1..1295

[illegible]

```

CDS
1. .1062
/ gene="CYP3A43/CYP3A4"
/ codon_start=1
/ product="cytochrome P450"
/ protein_id="CA91345.1"
/ db_xref="GI:30840241"
/ translation="MDLIPNFMETWLVATSLVLLYIVFGAYSMVDVITSTFGNID
SLANPODPFVNTKLLRFDLPFLIPILFVLEVLNLCVFPREVNTFLRKS
VKRMKSERLEDQKRVDFQLMIDSONSKETESHKALSDLELVAQSIIFIFAGYET
SSVLSTMYELATHPDVVOOKLOEIDAVLPNKAPPTDVLQWYLDVMVNETLRLFP
IAMLERVCCKVEILNGMFI PKGVVMIPSYALHRDPKYWTBEKELPERFSKKNKDN
IDPIIYTPGSGPRNCIMRPFALNMKLALIRVLQNFSPKPKETQIPLKLSLGGLLQ
PEKPVVLKVESRDGTVSGA"

ORIGIN
Query Match 2.5%; Score 48.4; DB 9; Length 1062;
Best Local Similarity 50.7%; Pred. No. 0.011;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

Qy 997 AAAAGCTGTCGATGAGAGGATCTCGCTCAGATCAGTAACTGTTATTTGCTGGATG 1056
Db 413 AAGCTCTGTCGATGAGGATCTCGCTGAGTAACTGTTATTTGCTGGATG 472

Qy 1057 AAATCTTTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTT 1116
Db 473 AAACACGAGCAGTGTCTCTCTCTCATTAATGATGACCTGCGCCTCAGCTGATGTC 532

Qy 1117 AGGATAAATCTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGCTAGACGAA 1176
Db 533 AGCAGAACTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGCTAGACGAA 1176

Qy 1177 TTAATGCGTTC-----CTTATCTCGAAGGTTTGTAAAGAGTCTCTTCGCTAGACC 1230
Db 593 ATACTGTCTACAGATGAGTATCTTACATGTTGTTGATGATGAAACGCTCAGATTATTC 652

Qy 1231 CTCCTAGTCCGTATGCTAACCGTGAATGCTTAAAGAGTGAAGACTTCA 1278
Db 653 CAATTGCTATGAGACTTGAGAGGCTTGCAAAAAGATGTTGAGATCA 700

RESULT 10
CQ717693 1197 bp DNA linear PAT 03-FEB-2004
LOCUS
DEFINITION
Sequence 3627 from Patent WO02068579.
ACCESSION
CQ717693
VERSION
CQ717693.1 GI:42278550
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Xtla, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL
Patent: WO 02068579-A 3627 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. .1197
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 2.5%; Score 48.4; DB 6; Length 1197;
Best Local Similarity 50.7%; Pred. No. 0.011;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

Qy 997 AAAAGCTGTCGATGAGAGGATCTCGCTCAGATCAGTAACTGTTATTTGCTGGATG 1056
Db 644 AAGCTCTGTCGATGAGGATCTGCGGCTGAGTAACTGTTATTTGCTGGATG 703

Qy 1057 AAATCTTTCGACAGTCTTGACATGGATGTTTACCGACTCTCAGAGACAAAGCCGTT 1116
Db 704 AAACACGAGCAGTGTCTCTCTCTCATTAATGATGACCTGCGCCTCAGCTGATGTC 763

Qy 1117 AGGATAAATCTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGCTAGACGAA 1176
Db 764 AGCAGAACTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGCTAGACGAA 1176

Qy 1177 TTAATGCGTTC-----CTTATCTCGAAGGTTTGTAAAGAGTCTCTTCGCTAGACC 1230
Db 824 ATACTGTCTACAGATGAGTATCTTACATGTTGTTGATGATGAAACGCTCAGATTATTC 883

Qy 1231 CTCCTAGTCCGTATGCTAACCGTGAATGCTTAAAGAGTGAAGACTTCA 1278
Db 884 CAATTGCTATGAGACTTGAGAGGCTTGCAAAAAGATGTTGAGATCA 931

RESULT 11
HSA563376 1365 bp mRNA linear PRI 21-MAY-2003
LOCUS
DEFINITION
Homo sapiens partial mRNA for cytochrome P450 (CYP3A43 and CYP3A4
genes).
ACCESSION
AJ563376
VERSION
AJ563376.1 GI:30840238
KEYWORDS
CYP3A4 gene; CYP3A43 gene; cytochrome P450.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
Finta, C. and Zaphiropoulos, P.G.
Intergenic mRNA molecules resulting from trans-splicing
J. Biol. Chem. 277 (8), 5882-5890 (2002)
2
MEDLINE
21839017
PUBMED
11726664
2 (bases 1 to 1365)
Zaphiropoulos, P.G.
Direct Submission
TITLE
Submitted (14-MAY-2003) Zaphiropoulos P.G., Bioscience, Karolinska
Institute, Novum, 141 57 Huddinge, SWEDEN
JOURNAL
FEATURES
source
1. .1365
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="liver"
1. .71
/ gene="CYP3A43"
1. .>71
/ gene="CYP3A43"
/ codon_start=1
/ product="cytochrome P450"
/ protein_id="CAD91644.1"
/ db_xref="GI:31043862"
/ translation="MDLIPNFMETWLVATSLVLLY"
72. .1365
/ gene="CYP3A4"
<72. .1365
/ gene="CYP3A4"
/ codon_start=2
/ product="cytochrome P450"
/ protein_id="CAD91645.1"
/ db_xref="GI:31043863"
/ translation="FYDQQQVLAITDPDMIKTVLKVCYSVFTNRPRFPVGFVGMKSA
ISAEDDEWKKRLSLLSPETSGKKNVPIIAQGVLVNLRREAGTGPVTLKOV
FGAYSMVDVITSTFGNIDSLNNPDPEVNTKKLRFDLPFLIPILFVLEVLNLCVFPREVNTFLRKS
EVLNLCVFPREVNTFLRKSVMKESRLEDQKRVDFQLMIDSONSKETESHKALSDLELVAQSIIFIFAGYET
DLELVAQSIIFIFAGYETSSVLSTMYELATHPDVVOOKLOEIDAVLPNKAPPTDVLQWYLDVMVNETLRLFP
IAMLERVCCKVEILNGMFI PKGVVMIPSYALHRDPKYWTBEKELPERFSKKNKDNIDPIIYTPGSGPRNCIMRPFALNMKLALIRVLQNFSPKPKETQIPLKLSLGGLLQ
WTEPEKFLPERFSKKNKDNIDPIIYTPGSGPRNCIMRPFALNMKLALIRVLQNFSPKPKETQIPLKLSLGGLLQ
KPKETQIPLKLSLGGLLQPEKPVVLKVESRDGTVSGA"

ORIGIN

```



```

Query Match      2.5%; Score 48.4; DB 9; Length 1365;
Best Local Similarity 50.7%; Pred. No. 0.012;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTTCGGATCGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTGCTGATATG 1056
DB 716 AAGCTCTTCGGATCGAGCTCGTGGCCCAATCAATATATCTTTATTTTCTGCTATG 775
QY 1057 AAACTTCTTCGACAGCTTTGACATGATGTTTCCCGACTCTCAGAAGACAAAGCCGTTTC 1116
DB 776 AAACCCAGCAGAGTGTCTCTCTTCATATGATGTAAGTGGCCACTCACCTCATGTTCC 835
QY 1117 AGGATAAACTTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGGTAGACGAAC 1176
DB 836 AGCAGAAAATCGAGAGGAAATTCATGACGATTTTACCCAAATAGGCCACCAACCTCATG 895
QY 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGAGTCTTCTGCTAGACC 1230
DB 896 ATACTGTGCTACAGATGAGATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 955
QY 1231 CTCCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
DB 956 CAATTGCTATGAGACTTGAGAGGGTCTGCAGAAAAGATGTTGAGATCA 1003

RESULT 12
LOCUS      CQ812569
DEFINITION Sequence 1 from Patent WO2004038015.
ACCESSION CQ812569
VERSION    CQ812569.1 GI:47602049
KEYWORDS   synthetic construct
SOURCE     synthetic construct
ORGANISM   artificial sequences.
REFERENCE  1
AUTHORS    Tickle, I.J., Vorheine, C., Williams, P.A., Jhoti, H. and Kirtan, S.B.
TITLE      Crystall. structure of cytochrome p450 3a4 and its use
JOURNAL    Patent: WO 2004038015-A 1 06-MAY-2004;
Astex Technology Limited (GB)
FEATURES   Location/Qualifiers
            source
            1..1458
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="Codes for SEQ ID NO: 2, a human 3A4 P450 protein
            truncated in its N-terminal region and the region is
            replaced by a short N-terminal sequence. The protein also
            comprises a poly- histidine tag at the C-terminus"

ORIGIN
Query Match      2.5%; Score 48.4; DB 6; Length 1458;
Best Local Similarity 50.7%; Pred. No. 0.012;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTTCGGATCGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTGCTGATATG 1056
DB 716 AAGCTCTTCGGATCGAGCTCGTGGCCCAATCAATATATCTTTATTTTCTGCTATG 856
QY 1057 AAACTTCTTCGACAGCTTTGACATGATGTTTCCCGACTCTCAGAAGACAAAGCCGTTTC 1116
DB 836 AAACCCAGCAGAGTGTCTCTCTTCATATGATGTAAGTGGCCACTCACCTCATGTTCC 916
QY 1117 AGGATAAACTTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGGTAGACGAAC 1176
DB 896 ATACTGTGCTACAGATGAGATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 976
QY 1231 CTCCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
DB 956 CAATTGCTATGAGACTTGAGAGGGTCTGCAGAAAAGATGTTGAGATCA 1084

RESULT 13
LOCUS      AX960331
DEFINITION Sequence 7 from Patent WO03102192.
ACCESSION AX960331
VERSION    AX960331.1 GI:40880462
KEYWORDS   synthetic construct
SOURCE     synthetic construct
ORGANISM   artificial sequences.
REFERENCE  1
AUTHORS    Cosme, J., Ward, A., Vuillard, L., Williams, P. and Hamilton, B.
TITLE      Methods of purification of cytochrome p450 proteins and of their
            crystallizing
JOURNAL    Patent: WO 03102192-A 7 11-DEC-2003;
Astex Technology Limited (GB)
FEATURES   Location/Qualifiers
            source
            1..1458
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="Description of Artificial Sequence: 3A4"

ORIGIN
Query Match      2.5%; Score 48.4; DB 6; Length 1458;
Best Local Similarity 50.7%; Pred. No. 0.012;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTTCGGATCGAGAGGTACTCGCTCAGATCAGTAACCTGTTATTTCTGCTGATATG 1056
DB 716 AAGCTCTTCGGATCGAGCTCGTGGCCCAATCAATATATCTTTATTTTCTGCTATG 856
QY 1057 AAACTTCTTCGACAGCTTTGACATGATGTTTCCCGACTCTCAGAAGACAAAGCCGTTTC 1116
DB 836 AAACCCAGCAGAGTGTCTCTCTTCATATGATGTAAGTGGCCACTCACCTCATGTTCC 916
QY 1117 AGGATAAACTTCGAGAGAAATTTGTTCAGATCGACACGGATATGCTTACGGTAGACGAAC 1176
DB 896 ATACTGTGCTACAGATGAGATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 976
QY 1231 CTCCTAGTCCGATGCTAACCGTGAATGCTTAAAGGATGAAGACTTCA 1278
DB 1037 CAATTGCTATGAGACTTGAGAGGGTCTGCAGAAAAGATGTTGAGATCA 1084

RESULT 14
LOCUS      CQ846981
DEFINITION Sequence 52 from Patent WO2004025244.
ACCESSION CQ846981
VERSION    CQ846981.1 GI:50896099
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS    Boulell, J.M., Godber, B.L., Hart, D.J., Bockett, N.A. and
            Kozlowski, R.S.
TITLE      Protein arrays and uses thereof
JOURNAL    Patent: WO 2004025244-A 52 25-MAR-2004;
SENSE PROTEOMIC LIMITED (GB)
FEATURES   Location/Qualifiers
            source
            1..1512
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"

```

```

/db_xref="taxon:9606"

ORIGIN
Query Match          2.5%; Score 48.4; DB 6; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.012;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTCCGATGAGGAGGTACTCGCTCAGATCAGTAACCTGCTTATTTGCTGGATATG 1056
DB 863 AAGCTCTGTCGATCTGGAGCTCGTGCCCAATCAATTTATCTTTTGGCTGCTATG 922

QY 1057 AAATCTTCGACAGTCTTGACATGGATGTTTCCCGACTCTCAGAAGACAAAGCCGTTT 1116
DB 923 AAACACGAGCAGTGTCTCTCTTCATATGATGTAAGTGGCCACTCACCCTGATGTC 982

QY 1117 AGGATAAACTTCGAGAAGAAATTTGTTCAGATCGACACGGATATGCTTACGCTAGACGAAC 1176
DB 983 AGCAGAACTCGAGGAGGAATTTGATGCAATTTTACCCCAATAAGGCACACCCACCTATG 1042

QY 1057 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGATGAAAGACTTCA 1278
DB 1177 TTAATGCGTTGC-----CTTATCTCGAAGCGTTTGTAAAGGATGAAAGACTTCA 1230

QY 1043 ATACTGTCTACAGATGAGTATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 1102
DB 1043 ATACTGTCTACAGATGAGTATCTTGACATGGTGGTGAATGAACGCTCAGATTATTC 1102

QY 1231 CTCCTAGTCCGTATGCTTAACCGTGAATGCTTAAAGGATGAAAGACTTCA 1278
DB 1103 CAATTGCTATGAGACTTGAGAGGGTCTGCAAAAAGATGTTGAGATCA 1150

Search completed: December 14, 2004, 09:31:00
Job time : 8085.28 secs

RESULT 15
E10636 Human cdna encoding cytochrome P4503A4. linear PAT 29-SEP-1997
DEFINITION E10636
ACCESSION E10636
VERSION E10636.1 GI:22027692
KEYWORDS JP 1996027197-A/5.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
TITLE ANTI-BODY RECOGNIZING CYTOCHROME P4502A4 ORIGINATED FROM MAN
JOURNAL Patent: JP 1996027197-A 5 30-JAN-1996;
SUMITOMO CHEM CO LTD
COMMENT OS Homo sapiens (human)
PN JP 1996027197-A/5
PD 30-JAN-1996
PF 13-JUL-1994 JP 1994161552
PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
PI YABUSAKI YOSHIYASU
PC C07K16/18, C12N15/09, C12N1/19, C12N9/02, G01N33/53, (C12N1/19, PC
C12R1:865);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1512
FT /organism="Homo sapiens"
FT /tissue_type="liver"
FT CDS 1..1512
FT /product="cytochrome P4503A4".
FT Location/Qualifiers
FT source 1..1512
FT /organism="Homo sapiens"
FT /mol_type="genomic RNA"
FT /db_xref="taxon:9606"

ORIGIN
Query Match          2.5%; Score 48.4; DB 6; Length 1512;
Best Local Similarity 50.7%; Pred. No. 0.012;
Matches 146; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 997 AAAAGCTCCGATGAGGAGGTACTCGCTCAGATCAGTAACCTGCTTATTTGCTGGATATG 1056

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2004, 21:43:25 ; Search time 412 Seconds  
(without alignments)  
482.884 Million cell updates/sec

Title: US-10-066-007A-1  
Perfect score: 2852  
Sequence: 1 MFILVLTGALGLAFAFSWAS.....RIVGREXGYQMRLOVXPVE 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1595576 seqs, 357178320 residues

Total number of hits satisfying chosen parameters: 1585576

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2852	100.0	557	14	US-10-066-007-1
2	2852	100.0	557	14	Sequence 1, Appli
3	399	14.0	504	17	Sequence 3, Appli
4	398.5	14.0	503	15	US-10-804-772-24
5	398.5	14.0	503	15	Sequence 24, Appl
6	394	13.8	485	17	US-10-313-963A-56
7	392.5	13.8	543	17	Sequence 56, Appl
8	387	13.6	517	16	US-10-146-575-2
9	387	13.6	523	16	US-10-690-991-2
10	385.5	13.5	503	9	US-10-425-115-294241
11	384.5	13.5	513	16	US-10-437-963-110700
12	384.5	13.5	650	16	US-10-437-963-157051
13	384	13.5	537	15	US-10-686-947-272
					Sequence 272, App
					Sequence 164419, A
					Sequence 38180, A

14	382	13.4	526	17	US-10-425-115-360975
15	381.5	13.4	527	17	US-10-425-115-233616
16	381	13.4	520	16	US-10-437-963-189675
17	380	13.3	529	17	US-10-425-115-209345
18	377.5	13.2	523	17	US-10-425-115-191779
19	377.5	13.2	527	15	US-10-425-114-67055
20	376.5	13.2	547	15	US-10-425-114-65616
21	374.5	13.1	461	16	US-10-686-947-176
22	374.5	13.1	528	17	US-10-425-115-224659
23	372.5	13.0	516	16	US-10-437-963-169524
24	370	13.0	521	16	US-10-686-947-216
25	368.5	12.9	523	16	US-10-437-963-184472
26	368.5	12.9	547	15	US-10-425-114-61218
27	368.5	12.9	560	15	US-10-425-114-59349
28	368.5	12.9	562	15	US-10-425-114-59350
29	367	12.9	520	9	US-09-992-901-2
30	367	12.9	521	16	US-10-686-947-264
31	367	12.9	525	16	US-10-437-963-174346
32	365.5	12.8	501	17	US-10-425-115-244178
33	365.5	12.8	662	15	US-10-425-114-38827
34	365	12.8	520	15	US-10-114-270-60
35	364	12.7	408	16	US-10-437-963-191588
36	363.5	12.7	512	16	US-10-432-609-5
37	359.5	12.6	506	16	US-10-437-963-196184
38	357.5	12.5	541	17	US-10-425-115-34446
39	354	12.4	516	16	US-10-437-963-190896
40	353.5	12.4	525	17	US-10-425-115-223573
41	353.5	12.4	528	16	US-10-437-963-150309
42	353	12.4	524	15	US-10-424-599-199859
43	352.5	12.4	540	16	US-10-437-963-118572
44	352	12.3	533	15	US-10-424-599-269344
45	351	12.3	512	16	US-10-432-609-6

ALIGNMENTS

RESULT 1  
US-10-066-007-1  
; Sequence 1, Application US/10066007  
; Publication No. US20030077691A1  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/10/066,007  
; CURRENT FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US/09/518,386  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: TRANSIT  
; LOCATION: (1)...(26)  
US-10-066-007-1

Query Match 100.0%; Score 2852; DB 14; Length 557;  
Best Local Similarity 100.0%; Pred. No. 6.8e-231;  
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFILVLTGALGLAFAFSWASIAFFSLYLAPRRSLYLNQGNHNYFTGNFLDILSARTG 60  
DB 1 MFILVLTGALGLAFAFSWASIAFFSLYLAPRRSLYLNQGNHNYFTGNFLDILSARTG 60

app 1: cont

QY 61 BEHAKYREKYSTLRFAAGIAGAPVLNSTDPKVFHVMKEAYDYPKPGMAARVLRATG 120  
Db 61 BEHAKYREKYSTLRFAAGIAGAPVLNSTDPKVFHVMKEAYDYPKPGMAARVLRATG 120  
QY 121 VVTAEGEAHKEHRRIMIPSLSAQAVKSMWPIFLEKGMELVDKMWEDAAEKDMVAGSAGE 180  
Db 121 VVTAEGEAHKEHRRIMIPSLSAQAVKSMWPIFLEKGMELVDKMWEDAAEKDMVAGSAGE 180  
QY 181 KKAIRLETEGVVDKDWGRATLDVMAAGFDYKSDSLQNKTNELVYAFVGLTDCGFAPTL 240  
Db 181 KKAIRLETEGVVDKDWGRATLDVMAAGFDYKSDSLQNKTNELVYAFVGLTDCGFAPTL 240  
QY 241 SFKAIMWDFVYFRTMKRRHEIPTQGLAVSRVYGIEMEQKQAVLGASDAQVKKDV 300  
Db 241 SFKAIMWDFVYFRTMKRRHEIPTQGLAVSRVYGIEMEQKQAVLGASDAQVKKDV 300  
QY 301 QGRDILSLVLRANTANLPSQKLSDEEVLQAIQSNLLFAGYETSSVLTWVHRLSDEKA 360  
Db 301 QGRDILSLVLRANTANLPSQKLSDEEVLQAIQSNLLFAGYETSSVLTWVHRLSDEKA 360  
QY 361 VQDKLREBEICQIDTDMPTLDELNALPYLEAFVKSLSRLDPPSPYANRECLKDEDFIPLAE 420  
Db 361 VQDKLREBEICQIDTDMPTLDELNALPYLEAFVKSLSRLDPPSPYANRECLKDEDFIPLAE 420  
QY 421 PVIGRDSGVINEVRITKGTWMLPLFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA 480  
Db 421 PVIGRDSGVINEVRITKGTWMLPLFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA 480  
QY 481 PYGHOASFIISGPRACFGWRFAVAKMFLVTLRRVQFEPIIISHPEYEHITLIISRPRI 540  
Db 481 PYGHOASFIISGPRACFGWRFAVAKMFLVTLRRVQFEPIIISHPEYEHITLIISRPRI 540  
QY 541 GREKEGYQMRLOQKXPE 557  
Db 541 GREKEGYQMRLOQKXPE 557

RESULT 2  
US-10-066-007-3  
; Sequence 3, Application US/10066007  
; Publication No. US20030077691A1  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/10/066,007  
; CURRENT FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US/09/518,386  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104668.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Phaffia rhodozyma  
US-10-066-007-3

Query Match 100.0%; Score 2852; DB 14; Length 557;  
Best Local Similarity 100.0%; Pred. No. 6.8e-231;  
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MFILVLLTGALGAFAFSWASIAFFSLYLAPRRSSLYNLOQPNHTNYFTGNFILDILSARTG 60  
Db 1 MFILVLLTGALGAFAFSWASIAFFSLYLAPRRSSLYNLOQPNHTNYFTGNFILDILSARTG 60  
QY 61 BEHAKYREKYSTLRFAAGIAGAPVLNSTDPKVFHVMKEAYDYPKPGMAARVLRATG 120

Db 61 BEHAKYREKYSTLRFAAGIAGAPVLNSTDPKVFHVMKEAYDYPKPGMAARVLRATG 120  
QY 121 VVTAEGEAHKEHRRIMIPSLSAQAVKSMWPIFLEKGMELVDKMWEDAAEKDMVAGSAGE 180  
Db 121 VVTAEGEAHKEHRRIMIPSLSAQAVKSMWPIFLEKGMELVDKMWEDAAEKDMVAGSAGE 180  
QY 181 KKAIRLETEGVVDKDWGRATLDVMAAGFDYKSDSLQNKTNELVYAFVGLTDCGFAPTL 240  
Db 181 KKAIRLETEGVVDKDWGRATLDVMAAGFDYKSDSLQNKTNELVYAFVGLTDCGFAPTL 240  
QY 241 SFKAIMWDFVYFRTMKRRHEIPTQGLAVSRVYGIEMEQKQAVLGASDAQVKKDV 300  
Db 241 SFKAIMWDFVYFRTMKRRHEIPTQGLAVSRVYGIEMEQKQAVLGASDAQVKKDV 300  
QY 301 QGRDILSLVLRANTANLPSQKLSDEEVLQAIQSNLLFAGYETSSVLTWVHRLSDEKA 360  
Db 301 QGRDILSLVLRANTANLPSQKLSDEEVLQAIQSNLLFAGYETSSVLTWVHRLSDEKA 360  
QY 361 VQDKLREBEICQIDTDMPTLDELNALPYLEAFVKSLSRLDPPSPYANRECLKDEDFIPLAE 420  
Db 361 VQDKLREBEICQIDTDMPTLDELNALPYLEAFVKSLSRLDPPSPYANRECLKDEDFIPLAE 420  
QY 421 PVIGRDSGVINEVRITKGTWMLPLFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA 480  
Db 421 PVIGRDSGVINEVRITKGTWMLPLFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA 480  
QY 481 PYGHOASFIISGPRACFGWRFAVAKMFLVTLRRVQFEPIIISHPEYEHITLIISRPRI 540  
Db 481 PYGHOASFIISGPRACFGWRFAVAKMFLVTLRRVQFEPIIISHPEYEHITLIISRPRI 540  
QY 541 GREKEGYQMRLOQKXPE 557  
Db 541 GREKEGYQMRLOQKXPE 557

RESULT 3  
US-10-804-772-24  
; Sequence 24, Application US/10804772  
; Publication No. US20040244077A1  
; GENERAL INFORMATION:  
; APPLICANT: Azpiroz, Ricardo  
; APPLICANT: Choe, Sungkwa  
; APPLICANT: Feldmann, Kenneth A.  
; TITLE OF INVENTION: DWP4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 11696-070001  
; CURRENT APPLICATION NUMBER: US/10/804,772  
; CURRENT FILING DATE: 2004-03-18  
; PRIOR APPLICATION NUMBER: US/09/502,426  
; PRIOR FILING DATE: 2000-02-11  
; PRIOR APPLICATION NUMBER: US 60/119,657  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: US 60/119,658  
; PRIOR FILING DATE: 1999-02-11  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-804-772-24

Query Match 14.0%; Score 399; DB 17; Length 504;  
Best Local Similarity 28.4%; Pred. No. 1.4e-24;  
Matches 156; Conservative 83; Mismatches 182; Indels 128; Gaps 24;  
QY 13 LAAFSW--ASIAFFSLYLAPRRS-SLYNLOQ-PNHTNY-FTGNFILDILSARTG-----E 61  
Db 7 LAMETWLLLAWSLVLLLYGTHSHGLFKKLGIPGPTPLPFLGN---ILSYHKGFCMFDM 63  
QY 62 BEHAKYREKYSTLRFAAGIAGAPVLNSTDPKVFHVMKEAYDYPKPGMAARVLR 113  
Db 64 CHKKYKQWGF---YDG--QQPVLAITDPMKLVLVKECYSVFTNRPFPVGFWMKSAI 118

QY 114 RIATGGVVTAGEAHRHRRIMFSLSAQAVKSMVPIFLEKGMELVDKQMEDAAEKDMA 173  
DB 119 SI-----ABDEEWKRLRLSPTFTSGKLKEMVPIIAQYGDVLYVNLRLREA----- 163  
QY 174 VCESAGEKATLETETEGVDVQKMGVGRATLDVYALAGFDYKSDSLQNKTNELYVAVGLTD 233  
DB 164-----RETGKPVTLKDVFGAYSDMTSTSSFGVNDLNNPD----- 201  
QY 234 GPAPLDSFKAIM-WDFV-PYFRTMK-RRHEIPTQGLV-----AVSRVVG 275  
DB 202---PLVENTKLLRDFLDPPFLSITVPFPILEVLNLCVFPREVTNPLKAVKMKKE 258  
QY 276 IELMEQKQAVLGSASDAQVDKQVQGRDILSLVRANIAANLPSQKLSDEEVLAQIS 334  
DB 259 SLEBDTQKRV-----DFQLMIDSHKSKETESHKALSDELVAQSI 301  
QY 335 NLLFAGYETSTVLTWVPHRLSEDAKAVQKLEEEICQI--DTPMTLDELNALPYLEAFV 392  
DB 302 IFIAGYETTSVLSFIMVELATHPDVQKQLEEDAVLPNKAPTYDTVLQMEYLDV 361  
QY 393 KESLRLDPPSYANRECLKDEDFPLAEPVIGRDSVINEVRITKGTWVMLPLFNINRSK 452  
DB 362 NETLFLFIAMRLERLVCKKDE-----INGMFIPKGVVWMPISVALHRDP 406  
QY 453 FIYGEDAEFRPERLEDDVDSLNS-IEAPYGHQASFTSGPRACFGWFAVAEMKALFV 511  
DB 407 -KYWTEPEKFLPERFSKXKNIDPIYTPFG-----SGPRNCIGMRFALNMKALIR 459  
QY 512 TLRRVQFEP 520  
DB 460 VLQNFSEKP 468

## RESULT 4

US-10-313-963A-56  
; Sequence 56, Application US/10313963A  
; Publication No. US20040002078A1  
; GENERAL INFORMATION:  
; APPLICANT: Boutell, Jonathan  
; APPLICANT: Godber, Benjamin  
; APPLICANT: Hart, Darren  
; APPLICANT: Blackburn, Jonathan  
; TITLE OF INVENTION: Arrays  
; FILE REFERENCE: KIL-001  
; CURRENT APPLICATION NUMBER: US/10/313,963A  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: US 60/335,806  
; PRIOR FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: US 60/410,815  
; PRIOR FILING DATE: 2002-09-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-313-963A-56

Query Match 14.0%; Score 398.5; DB 15; Length 503;  
Best Local Similarity 28.7%; Pred. No. 1.5e-24;  
Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;

QY 13 LAAPSW--ASTAFPSLYLAPRRS-SLYNLOG-PNHTNY-FTGNFLDILSARTG-----E 61  
DB 7 LAMETWLLAVSLVLLYLYGTHSHGLPKLGIPOPTPLPLGN---ILSVHKGFCMFDM 63  
QY 62 EHAKYREKYGSTLRFAAGAPVLNSTDPKVFNVHM-KEAYDY---PKPGVAARVLRAT 117  
DB 64 CHKYGKVGWF---YDG--QQPVLAITDPMIKTVLVKECYSVFTNRRPFGVGMKSA- 117  
QY 118 GDGVVTAAGEAHRHRRIMFSLSAQAVKSMVPIFLEKGMELVDKQMEDAAEKDMAVGS 177

DB 118 ---ISIADEEWKRLRLSPTFTSGKLKEMVPIIAQYGDVLYVNLRLREA----- 164  
QY 178 AGEKKAATLET-EGVDVQKMGVGRATLDVYALAGFDYKSDSLQNKTNELYVAVGLTDGFA 236  
DB 165-----BTGKPVTLKDVFGAYSDMTSTSSFGVNDLNNPD----- 201  
QY 237 PTLDSFKAIM-WDFV-PYFRTMK-RRHEIPTQGLAV---SRRVGTIELMEQKQAVLGS 290  
DB 202 PFVENTKLLRDFLDPPFLSITVPFPILEVLNLCVFPREVTNPLKAVKMKKE 256  
QY 291 SDOAVDKQVQGRDILSLVRANIAANLPSQKLSDEEVLAQISLNLFPAGYETTSVLTW 350  
DB 257 KESLRLDTPKRVDFLDPPFLSITVPFPILEVLNLCVFPREVTNPLKAVKMKKE 316  
QY 351 MFHRLSEDAKAVQKLEEEICQI--DTPMTLDELNALPYLEAFVKSLSRLDPPSYANRE 408  
DB 317 IMVELATHPDVQKQLEEDAVLPNKAPTYDTVLQMEYLDVQVNETLRLFFIARLERV 376  
QY 409 CLKDEDFPLAEPVIGRDSVINEVRITKGTWVMLPLFNINRSKFTYGSDEAEFRPERWL 468  
DB 377 CKKDE-----INGMFIPKGVVWMPISVALHRDP-KYWTEPEKFLPERFS 420  
QY 469 EDVTDLSNS-IEAPYGHQASFTSGPRACFGWFAVAEMKALFVTLRRVQFEP 520  
DB 421 KXKNONIDPIYTPFG-----SGPRNCIGMRFALNMKALIRVLQNFSEKP 467

## RESULT 5

US-10-146-575-2  
; Sequence 2, Application US/10146575  
; Publication No. US20030059600A1  
; GENERAL INFORMATION:  
; APPLICANT: Lichter, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/10/146,575  
; CURRENT FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: US/09/144,367  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: H. sapiens  
US-10-146-575-2

Query Match 13.9%; Score 396.5; DB 14; Length 503;  
Best Local Similarity 28.7%; Pred. No. 2.2e-24;  
Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;

QY 13 LAAPSW--ASTAFPSLYLAPRRS-SLYNLOG-PNHTNY-FTGNFLDILSARTG-----E 61  
DB 7 LAMETWLLAVSLVLLYLYGTHSHGLPKLGIPOPTPLPLGN---ILSVHKGFCMFDM 63  
QY 62 EHAKYREKYGSTLRFAAGAPVLNSTDPKVFNVHM-KEAYDY---PKPGVAARVLRAT 117  
DB 64 CHKYGKVGWF---YDG--QQPVLAITDPMIKTVLVKECYSVFTNRRPFGVGMKSA- 117  
QY 118 GDGVVTAAGEAHRHRRIMFSLSAQAVKSMVPIFLEKGMELVDKQMEDAAEKDMAVGS 177  
DB 118 ---ISIADEEWKRLRLSPTFTSGKLKEMVPIIAQYGDVLYVNLRLREA----- 164  
QY 178 AGEKKAATLET-EGVDVQKMGVGRATLDVYALAGFDYKSDSLQNKTNELYVAVGLTDGFA 236  
DB 165-----BTGKPVTLKDVFGAYSDMTSTSSFGVNDLNNPD----- 201  
QY 237 PTLDSFKAIM-WDFV-PYFRTMK-RRHEIPTQGLAV---SRRVGTIELMEQKQAVLGS 290  
DB 202 PFVENTKLLRDFLDPPFLSITVPFPILEVLNLCVFPREVTNPLKAVKMKKE 256  
QY 291 SDOAVDKQVQGRDILSLVRANIAANLPSQKLSDEEVLAQISLNLFPAGYETTSVLTW 350

Db 257 KEKRLSDTKQHRVDFQLQMDSONSKETSHKALSDLELVAQSIIFIFAGYETTSSVLSF 315  
 QY 351 MFHRLSEDAVQKLRREICQI--DTDMPTLDELNALPYLEAFVKESELRULPPGPPYANRE 408  
 Db 317 IYVELATHPDVQKQLEEDAVLPNKAPPTDYTVLQMEVLDVMVNETLRLPFIARLERV 376  
 QY 409 CLNDEDFIPLAEPVIGRDSVINEVRIITKGTWMLPLFNIRSKFIYGEDAEERPRRWL 468  
 Db 377 CKXKDE-----INGMFIKGMWVMIPEYALHRDP-KYWTSEKFLPERFS 420  
 QY 469 EDVTDLSNS-IEAPYGHQASFISSGRACFGMRFAVAEMKAFLEFVTLRRVQPEP 520  
 Db 421 KKKKNDIDPIYITPFG-----SGPRNCIGMFEALWMMKALIRVLQNFSPKP 467

## RESULT 6

US-10-690-991-2  
 ; Sequence 2, Application US/10690991  
 ; Publication No. US20040243319A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tickle, Ian J  
 ; APPLICANT: Vonthein, Clemens  
 ; APPLICANT: Williams, Pamela A  
 ; APPLICANT: Jhoti, Harren  
 ; APPLICANT: Kirtson, Stewart Brian  
 ; TITLE OF INVENTION: Crystal structure of cytochrome P450  
 ; FILE REFERENCE: 620-282  
 ; CURRENT APPLICATION NUMBER: US/10/690,991  
 ; PRIOR APPLICATION NUMBER: US 60/421,063  
 ; PRIOR FILING DATE: 2002-10-25  
 ; PRIOR FILING DATE: 2002-05-30  
 ; PRIOR FILING DATE: 2002-04-02  
 ; PRIOR FILING DATE: 2002-04-02  
 ; PRIOR FILING DATE: 2003-06-19  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 485  
 ; TYPE: PRT  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Human 3A4 P450 protein truncated in its N-terminal region to  
 ; OTHER INFORMATION: delete the hydrophobic trans-membrane domain, and the region  
 ; OTHER INFORMATION: replaced by a short N-terminal sequence.

## US-10-690-991-2

Query Match 13.8%; Score 394; DB 17; Length 485;  
 Best Local Similarity 28.7%; Pred. No. 3.4e-24;  
 Matches 142; Conservative 83; Mismatches 177; Indels 92; Gaps 20;  
 QY 47 FTGNFLDILSARTG-----EEHAKYREKYGSTLRFAGIAGAPVNLSTDPKVFNNHM-KE 99  
 Db 24 FTGN-----ILSYHKFGCMDFMECHKYKVGWP---YDG--QOPVLAITDPDMIKTVLKE 75  
 QY 100 AYDY---PKPGMAARVLIATGDGVVTAEGEAHKHRRIMIPSLSAQAVKSMVPFLEKG 156  
 Db 76 CYSVFTNRRPFPVPGVGMKSA-----ISIAEDEEWKRLRLSLLSPTTSGLKXEWPIIAQYG 131  
 QY 157 MELVDKMDAEAKDMVAGESAGEKKATELET-EGVDVKWVGRAITLDVNALAGFDYKSD 215  
 Db 132 DVLVRLRREA-----ETGKPVTLKQVFGAYSMDVITSTSFQVNIID 172  
 QY 216 SLQKNTNELYAVFGLTDPGAPFLDSFKAIM-WDFV-PYFRIMK-RREHIEPLTQGLAV-- 270  
 Db 173 SLNNPD-----PFVENTKKLRDFLDPFPLSITVFPFLPILEVLNICV 218  
 QY 271 -SRVVGIELEWQKQAVLGSDQAVDKVQGRDILSLVRAIANLPEKSLDSEV 329  
 Db 219 PFREVTNLRKSVKR-----MKESLEDTQKRVDFLQLMIDSONSKETSHKALSLEL 273

QY 330 LAQISNLLFAGYETSSVLTWTFHRLSEDAVQKLRREI CQI--DTDMPTLDELNALPY 387  
 Db 274 VAQSIIFIFAGYETTSSVLSFIMYELATHPDVQKQLEEDAVLPNKAPPTDYTVLQMEY 333  
 QY 388 LEAFYKESIRLDDPPSPYANRECLXDEDFIPLAEPVIGRDSVINEVRIITKGTWMLPLFN 447  
 Db 334 LDMVNETLRLPFIARLERVCKKDE-----INGMFIKGMWVMIPEYALHRDP-KYWTSEKFLPERFS 378  
 QY 448 INRSKFIYGEDAEERPRRWLEVDVTDLSNS-IEAPYGHQASFISSGRACFGMRFAVAEMK 506  
 Db 379 LHRDP-KYWTSEKFLPERFSKKNKNDIDPIYITPFG-----SGPRNCIGMFEALWMMKALIRVLQNFSPKP 431  
 QY 507 AELFVTLRRVQPEP 520  
 Db 432 LALIRVLQNFSPKP 445

## RESULT 7

US-10-425-115-294241  
 ; Sequence 294241, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 294241  
 ; LENGTH: 543  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(543)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_31433C.1.pep  
 ; US-10-425-115-294241

Query Match 13.8%; Score 392.5; DB 17; Length 543;  
 Best Local Similarity 26.1%; Pred. No. 5.4e-24;  
 Matches 150; Conservative 95; Mismatches 226; Indels 103; Gaps 23;  
 QY 6 LITGALGLAFAFWASIAFFSL----YLAPRRSLYLQGNHTNYFTGNFLDILSARTG- 60  
 Db 22 MVAATAAVALASWAFNALVHLVWRPVATRLRAQGVGPGPYT-FTTSGILGEIKRLRAG 80  
 QY 61 -----EEH-----AKYREKYGSTLRFAGIAGA-PVLNSTDPKVFNNHMKEAYD 102  
 Db 81 AAVTLVDVDDHDFIEMVQFHLRKIALYGRT--FVYTGARPNVCVADVNVVRVL---FD 135  
 QY 103 ----YPKPGMAARVLIATGDGVVTAEGEAHKHRRIMIPSLSAQAVKSMVPFLEKGM 158  
 Db 136 RTGLYKPNLINPHISRL-LGKGLVLTGDDWKHRRKVVHPAFNMDKLMATMTSDCTRS 194  
 QY 159 LVDRXOMEDAEAKDMVAGESAGEKKATELETGVDVVDKVGRAITLDVNALAGDYKSDSLQ 218  
 Db 195 MISSWDALQKQED-----QSGRGHVE---VELSSRFEELTADVISHTAP----GSSY 242  
 QY 219 NKTNELYAVFGLTDPGAPFLDSFKAIMMDVFPYFRIMKRRHEIPIQTGLAVSRVGIEL 278  
 Db 243 NEGKRVLAQRELGH-----IAFSTIF-----NVQIPALKYLTPEKMLRTRK 284  
 QY 279 MEQKQAVLGSDQAVDKKDV---QGRDILSLVRAIANLPEKSLDSEV 329  
 Db 285 LDRQVRLWMDIIEARLASDITAGGYNLDLGLMLEA--CASPPHHGEMWAPTILSMDEI 342

Query Match	13.6%;	Score 387;	DB 16;	Length 517;
Best Local Similarity	25.4%;	Pred. No. 1.5e-23;		
Matches 146;	Conservative 100;	Mismatches 212;	Indels 116;	Gaps 23;
QY	6	LLTGALGLAAESWASIAFFSLLYAP-----PRSSLYNLQGNHTNPTFTGNFLDLSARTGE	61	
Db	7	MYAAVAAVASWAFDAVVLWRPRAITRLRAQGVGGPGY-RPFSGNLGEI--RRLRD	63	
QY	62	BHA-----KYREKYGSLTRFAGIAGAPVLNSTDPKPVHNVKRAY	101	
Db	64	EGAGVULDVSSHDFPIVPVPHFRKWPVLYGKTTY-WFGAQPTICLADMSWVRQVLSRT	122	
QY	102	D-YPK-----PCYAARVLRITATGGVTAAGEAHKRHRRIIMPISLSAQAVKSGMPVIFLEKG	156	
Db	123	GWPKDVTNPYFAOLL-----GKGLVLTGDGEWKRHRKVHPAFNMDLKMMT-----	170	
QY	157	MELVDKMDAAEKDMAVGESACEKKATRLTEG---VDVKDWVGRATLDVMALAGFDY	212	
Db	171	-----VTMSDCTQSMISEWES-----ELGTHKGSABEILSQRTQEIATDVISHTA	216	
QY	213	KSDSLQNKTNELYYAVFGLTGDGPAPLTDGFKALIMDFVPY--FRMKRHRHPIPTQGLAV	270	
Db	217	--GSSYSEGKQVFLTAQREL-----QFLAFSMFLTI---QIPFRVLYPT	254	
QY	271	SRRYGIELMEOKKQAVLGASDCAVDKKOV--QGRDILSLIVRANIAANLPESQ-KLSDE	327	
Db	255	KKNKILSLDKVMSMTNIIKIRLTNNKQVTDGYGNDLLGLMLUE-TCAPHEGESQPQLSMD	313	
QY	328	EVLQAQINLLFAGYETSGTSLVTWVHRLSEDKAVODKLRFEETCOIDTD-MPTLDELNALP	386	

Query Match	13.6%	Score 387	DB 16	Length 523
Best Local Similarity	24.8%	Pred. No. 1.5e-23		
Matches 143	Conservative 94	Mismatches 245	Indels 94	Gaps 19
QY	3	ILVLITGAL---	GLAASFWSASIAFFSLYLAPRRSSLYNLOQGNHNTNYFTGFLDILSART	59
DB	10	ILVLVGRLLTSALVHLLWKPYAITKLFRG----	QGITGPKY-RLFVGSLEPEIKRMKA	62
QY	60	G-----	EBHAKVREKYGST-LRFAGIAGAPVLNSTDPKVFNHVMK	98
DB	63	AAAADAVAAGAHSHDFPIVLPQHSKWATDHGKTFLYNLGAVPAVSLGRVE-CVLRQVLEL	121	
QY	99	EAYDYPKPGMAARVLRIATGDGVVTAEGEAHKHRRIMI	PSLSAQAVKSMWPIFLKGMW	158
DB	122	RTGSFTKNYNNAN-LEALLGKGIILANGEDWEHRKVVHPAFNHDXLKPMFSVMVAESVES	180	
QY	159	LVDKMMEDAAEKMANAGESAGEKKATRELETEGVDDVKDWYGRATLDVMALAGEDYKSDSLQ	218	
DB	181	MVQRWQSQI-----	QQAGNNQVE-----LDLSRELSELTSDDVITRSAF-----GSSH	222
QY	219	NKTMELVAVFGLTD-GFAPTLDSFKALMWDVPVPTMTKRHEHPLTQGLAVSRVIGIE	277	
DB	223	EEGKEVYQAOKEQELAFSSLDVPAVFLRKLPIIRGNTRAHQL-----	VKKSRTMLME	277
QY	278	LMEQKQAVLGASDAQVDKKVQVGRDIIISLVIRANIAANLPESQKLSDEEVLQAIGISNLL	337	
DB	278	IIEGRLAKVEAAEAG-----	YGSDDLGLMLERALEREGLNGLVLTQIIHIDECKTFF	329
QY	338	FAGVETGSTVLTMFHRLSLSDKXAVQDKREEICQIDTDD-MPTLDLBNALPYLEAFVKESL	396	



Db 494 PRSITVQPQ 503

# RESULT 12

US-10-437-963-164419  
; Sequence 164419, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 164419  
; LENGTH: 650

; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_6331C.1.pap  
US-10-437-963-164419

Query Match 13.5%; Score 384.5; DB 16; Length 650;  
Best Local Similarity 27.1%; Pred. No. 3.3e-23;  
Matches 140; Conservative 78; Mismatches 208; Indels 91; Gaps 18;

QY 53 DILSARTGEBHAXREKXGTLRAGIAGAPVNLSTDPKVNVMKEAY-DYKPGMAAR 111

Db 208 DIVRVPFVHRNVRDN-GRPCCFSPGPIPTIDPAQVRDVLNKLGHFKPLPAL 266

QY 112 VLRIATGDGVVTAEGEAHRRIMIPSLSAQAVKSMVPIFLEKGMELDKMEDAEXD 171

Db 267 TKULA--DGLTSHDGEKWKVHRIMPAPFHEKLEKMLPAFSTCCBELVGKM----- 317

QY 172 MAVGESAGEKATRLTEGVVDKVGRAFLDVMALGF--DYKSD-----SLQKTNELY 225

Db 318 ----DSLGPDGSCSEL-----DVPENQSLTGDVISRTAFGSSYSEGRIFQLQTEQAELEF 368

QY 226 VAFVGLTDGFAPLDSFKAIMWDFVP--YFRTMKRHEIPLTQGLAVSRVGVIELMEQK 282

Db 369 IG-----ALQKVFPGMYLPTKKRR-----MRRINSE-VESI 401

QY 283 KQAVLGSASQAVDKVQGRDILSLIVRANIA---ANLPESQKLSDEEVLQISNLLFA 339

Db 402 LRGIIGKRM-QAIAEGESTNDLLGLLLESNMRHADENGRSSPGMTTEDIIECKLFYFA 460

QY 340 GYTSSVLTWMPHRLSEDKAVQDKLREEICQI-DTDMPTLDLNLALPYLEAFVRESRL 398

Db 461 GNETTSVLLTWVWLSMHPWQDRAREEVLGFGDKPEYEGLSRLKTKVTWVLYEVLRL 520

QY 399 DPSPYANRECLKDEDFIPLAEPVIGRDGVSINEVRIITKTMVMLPLFNINRSKFTYGED 458

Db 521 YPAIVFSRTYKEME-----IGVVYVPRGVILELPLFVLIHHDREIWDGD 565

QY 459 AEFERPERLWEDVDSLSNSTEA--PYGHOASFISSGRACFWRFPAVAKMKAFLVTLRRV 516

Db 566 VHEFRPERFAEGSRASNDRGALPFGW-----GPRVCIGQNFALLEAKMALCMILQRF 619

QY 517 QFEPIIISHPEYEHTLIISRPRVIGREKEGYQWRLQV 553

Db 620 EFELAASYTHAPHTVMTL-----HPMEGAQWKLKM 649

# RESULT 13

US-10-425-114-38180

; Sequence 38180, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 38180

; LENGTH: 537  
; TYPE: PRT

; ORGANISM: Zea mays  
; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3608-017-E5\_FLI.pap  
US-10-425-114-38180

Query Match 13.5%; Score 384; DB 15; Length 537;

Best Local Similarity 27.4%; Pred. No. 2.8e-23;  
Matches 162; Conservative 79; Mismatches 230; Indels 120; Gaps 24;

QY 4 LVLLTGALGAAFSWASIAFFS-LYLAPRR-SSLYNLQGNHTNY--FTGNFLDILSART 59

Db 23 LLIVYVVLG-ALLLWKAARLLDLRWEPRLERALARAOGLRGTSYRFLTG--DLREYRR 78

QY 60 GEEHAKYRE-----KYGSTLRFAGIAGAPVNLSTDPKVNVMK 98

Db 79 SKEEAWAPLPLRCHDIAGHVEPFHGAVLEHGKTC-FTWFGPVRVTVPDLDARDVMA 137

QY 99 EAY-DYKPGMAARVIRIATGDGVVTAEGEAHRRIMIPSLSAQAVKSMVPIFLEKGM 157

Db 138 NKGFHFERKPPA--LTKLPSDGVANHEGKWKVHRILNPAFHEKLEKMLPAFSACCE 195

QY 158 ELVDKMWEDAAEKDMVAGESAGEKATRLTEGVVDKVGRAFLDVMALGF-----D 211

Db 196 ELVSRW-----AQSLGPDGSCSEL-----DVPDELQTLTGDIVISRTAFGSSYLEG 239

QY 212 YKSDSLQKTNELVAVFVGLTDGFAPLDSFKAIMWDFVPVFTMKRHEIPLTQGLAVS 271

Db 240 RKIFQLQAEQAEERLMSII--DKFA-----VPGYMSLPTKNNRMQIKSEI 283

QY 272 RRVGIELMEQKQAVLGSASQAVDKVQGRDILSLIVRANIAANLPES----OKLSDE 327

Db 284 DSILRGLIGKRMQAMKQESD-----KD-----DLLGLLLESNARETGQSQPGQGLTME 334

QY 328 EVLAQISNLLFAGYETSTVLTWMPHRLSEDKAVQDKLREEICQI--DTDMPTLDLNLAL 385

Db 335 EVMECKLFYFAGNETTSVLLTWVWLSMHPWQDRAREEVLGFGKKNQPGYDGLSRL 394

QY 386 PYLEAFVRESRLDPSPYANRECLKDEDFIPLAEPVIGRDGVSINEVRIITKTMVMLPL 445

Db 395 KTVTMILYEVRLYPPAIFASRTYK-----VVVGVDVTPYAGVTLELPV 439

QY 446 PNINRSKFTYGEDAEERPERLWEDVDSLSNSTEA--PYGHOASFISSGRACFWRFPAVA 503

Db 440 LFIHDDPDIGSDAHEFRPERFAEGVARASKDRLAFFFGW-----GPRICIGQNFALL 493

QY 504 EMKAFLVTLRRVQFEPIIISHPEYEHTLIISRPR--IVGREKEGYQWRLQ 552

Db 494 EAKWALSMLQRFQFE---LAPTTHV-----PRRVMILRPPHMGAIKUR 535

# RESULT 14

US-10-425-115-360975

; Sequence 360975, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:



```

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 360975
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(526)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92390C.1.pep
US-10-425-115-360975

```

```

Query Match      13.4%; Score 382; DB 17; Length 526;
Best Local Similarity 23.3%; Pred. No. 4e-23;
Matches 137; Conservative 108; Mismatches 240; Indels 102; Gaps 20;

QY 1 MFILVLTGALGAFAWASIAFFSLYLAPRRSLYNLQPNHTYFTGNFLDIL-----S 56
Db 1 MVAVQVLAALLALLLWRLV--WRPHAVARSFARQGRPHYT-FLAGSLPEAIGLEA 57
QY 57 ARTG-----EEHAKYREKYGSTLRPAGIAGAPVLASTDPKVFNVHVKEA 100
Db 58 GRRGTAPLDAGCHDIMPVLLPQHRWVADYGRITLF-WIGPIPAVLSTDLQILKQVLTDR 116
QY 101 YDYPKPGMAARVLGIAGDGVTAEGEAKHRRIRIMPSLSAQAQVSMVPIFLEKGMELV 160
Db 117 TGLYKQDFMPLVFLFNGVILINGDDWGRHKVLPAPFNHTIISMNAV-----TAETV 172
QY 161 DKMEDAAEKDMVAGESAGEKATRLTEGVYKDWVGRATLDVMAAGFDYKSDSLQNK 220
Db 173 EQWQ-----QNRGQIHGSEESAEIDMHAENDLTSKNGRVAF-----G7SHRE 218
QY 221 TNELYVAFVGLTD-GFAPTLDSFKAIMWDFVPVFTMKRHEIPLTQGLAVSRVGIEM 279
Db 219 VEEVILMREKQKATATLDA--PILW-----YLPTRNLH-----VRLNKQLR 262
QY 280 EQKQAVLGASDAQVKKDVQ-----GRDILSLVRA--NIAANLPESQKLSDEVL 331
Db 263 SKIMSINQARLAADAGDRGGRGAVSGGDLGLLEAXRRSRNNGNGGTLTTTDEVID 322
QY 332 QISNLLFAGYETSTVLTWHLRSEKAVQDKLREEICQ-----IDTDMPTLDELNALP 386
Db 323 ECKTFFAAGQETTATLLWAMFLVAHPWQDKVREEVREFTCTSDGGEVPHADVLAKL 382
QY 387 YLEAFVKESSLRDPSPYANRECKLDEDFPLAEPVIGRDSGVINEVRITKGTWMLPLF 446
Db 383 LLYVVLLETSLYPIPIYIQRAND-----AVLGGIKVPQGITVISPIA 427
QY 447 NINRSKFIYGDABEEFRERWLEVDTSLSIEAPYGHQASPIGPRACFGWFAVEMK 506
Db 428 MLHRDKQWVGPDADDEFNREHGLTKAAKDKA-----LLSFLSLGPRVCTGQSGFIVEVQ 493
QY 507 AFLVTLRRVQFEPIIHPYEH-----ITLIISRPVIGREKEG 546
Db 484 VVMAMILRRFSF-----LSPKVTHPKVLLSLTPKLGMLPLV-RNVDG 526

```

## RESULT 15

```

US-10-425-115-233616
; Sequence 233616, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 233616
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_144651C.1.pep
US-10-425-115-233616

```

```

Query Match      13.4%; Score 381.5; DB 17; Length 527;
Best Local Similarity 24.5%; Pred. No. 4.4e-23;
Matches 138; Conservative 91; Mismatches 137; Indels 137; Gaps 23;

QY 38 LQGNHTNY--FTGNFLDILSARTGEE-----HAKYREKYGSTLR 75
Db 50 VQGLKGRYRLFTGDLRE--TARANREARKKPLPLGSHDIAPRVQPMHSHSTIKEYGR-LS 106
QY 76 PAGIAGAPVLNSTDPKVFNVHVKAEYDY--PKGMAARVLRIATGCVVTVABGEAKHRH 132
Db 107 FTWFGTPRVMIPOPELVKEVLSNKGFGKGRSNRIGRLI-----ANGLVNHGDEKAKH 162
QY 133 RRIMPSLSAQAQVSMVPIFLEKGMELVDKMDAEKMDAVGESAGEKATRLTEG-- 190
Db 153 RRLNPAFHHEKIKGMPVFSTCCIEITRW-----DNSMPSEGS 203
QY 191 -VDYKDWVGRATLDVMAAGF--DYKSD-----SLQNTNELYVAFVGLTGDGAPTLDSPK 243
Db 204 EIDWPEFQNLTDGVISRTAFGSNYQGRRIFELOGLAEELI-----Q 247
QY 244 AIMWDFV--VFRTMKRHEIPLTQGLAVSRVGIEMEQKQAVLGASDAQVKKDVQ 301
Db 248 SVQTFIPGYWFLPTKNRM-----RAIDVEIRKILRE-IICKR-----EKDTK 291
QY 302 GR-----DILSLVRA--NIAANLPESQKLSDEVLQISNLLFAGYETSTVLTWHLR 355
Db 292 NRETNDDLLGLLESNTQSGNAGSLGTLTDEVIIECKLFYFAGMETTSLVLTWTLVL 351
QY 356 SEDKAVQDKLREEI-CQIDTDMPTLDELNALPYLEAFVKESSLRDPSPYANRECKLDE 414
Db 352 SMHPEWQSRAREEVLSHFGRTTPDYDSLRLKVTWMLHEVLRILYPPATFLTRTYKEME 411
QY 415 FIPLAEPVIGRDSGVINEVRITKGTWMLPLFNINRSKFIYGEDAEFRERWLEVDTS 474
Db 412 -----LGGIKVPAGVDULLPVIFIHDPDINGKDSBFPNPERF----- 449
QY 475 LNSIEAPYGHQASFI---SGPRACFGWFAVEMKAFVTLRREVOE--PIIHPYEH 529
Db 450 ANGISSATRHQAAPFPFGGPRICIGOSFALLEARMTLCTILQRFSELSPSYTHAPTV 509
QY 530 ITLIISRPVIGREKEGYQMRIQ 552
Db 510 ITL-----HPQHGAIQLK 523

```

```

Search completed: December 14, 2004, 22:03:11
Job time : 414 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2004, 21:39:59 ; Search time 38 Seconds  
(without alignments)  
972.082 Million cell updates/sec

Title: US-10-066-007A-1

Perfect score: 2852

Sequence: 1 MFILVLLTGALGLAFAFSWAS.....RIVGREKEGYQMRLQVKE 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents AA.\*
- 2: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2852	100.0	557	3	US-09-518-386B-1
2	2852	100.0	557	3	US-09-518-386B-3
3	396.5	13.9	503	4	US-09-144-367-2
4	370	13.0	504	1	US-08-457-274A-25
5	370	13.0	504	5	PCT-US95-05758-25
6	367	12.9	520	4	US-09-527-073-2
7	345	12.1	524	4	US-09-976-594-533
8	338.5	11.9	503	4	US-09-583-447A-2
9	331	11.6	540	3	US-09-302-620B-99
10	330.5	11.6	507	1	US-08-457-274A-22
11	330.5	11.6	507	5	PCT-US95-05758-22
12	328	11.5	504	4	US-09-583-447A-4
13	328	11.5	540	3	US-09-302-620B-98
14	326.5	11.4	541	3	US-09-158-767-19
15	326.5	11.4	541	3	US-09-158-767-20
16	326.5	11.4	541	3	US-09-713-794-19
17	326.5	11.4	541	3	US-09-713-794-20
18	321	11.3	576	3	US-08-948-564-16
19	320.5	11.2	529	4	US-09-270-767-46468
20	318.5	11.2	526	1	US-08-298-426-4
21	315	11.0	1049	4	US-10-018-730A-4
22	294.5	10.3	489	4	US-09-852-067-4
23	294	10.3	522	3	US-09-302-620B-97
24	289	10.1	522	3	US-09-302-620B-96
25	286	10.0	548	4	US-09-248-796A-15537
26	278.5	9.8	507	4	US-09-248-796A-15540
27	276	9.7	420	4	US-09-583-447A-6

28	276	9.7	467	3	US-09-126-420A-17	Sequence 17, Appl
29	269.5	9.4	517	3	US-09-302-620B-100	Sequence 100, App
30	269.5	9.4	517	4	US-09-911-781-32	Sequence 32, Appl
31	269.5	9.4	517	4	US-10-400-902-32	Sequence 32, Appl
32	269	9.4	507	1	US-08-457-274A-23	Sequence 23, Appl
33	269	9.4	507	5	PCT-US95-05758-23	Sequence 23, Appl
34	267.5	9.4	517	3	US-09-302-620B-101	Sequence 101, Appl
35	263	9.2	523	3	US-09-302-620B-95	Sequence 95, Appl
36	259.5	9.1	510	3	US-08-948-564-4	Sequence 4, Appl
37	259	9.1	510	4	US-09-852-067-2	Sequence 2, Appl
38	257.5	9.0	509	4	US-10-140-002-108	Sequence 108, App
39	253.5	8.9	512	4	US-09-270-767-45399	Sequence 45399, A
40	251.5	8.8	508	3	US-09-126-420A-25	Sequence 25, Appl
41	246	8.6	512	3	US-09-302-620B-103	Sequence 103, App
42	243	8.5	512	3	US-09-302-620B-102	Sequence 102, App
43	241.5	8.5	476	1	US-08-313-075A-30	Sequence 30, Appl
44	240.5	8.4	485	4	US-09-248-796A-17581	Sequence 17581, A
45	239.5	8.4	1023	4	US-09-270-767-43827	Sequence 43827, A

#### ALIGNMENTS

RESULT 1  
US-09-518-386B-1  
; Sequence 1, Application US/09518386B  
; Patent No. 6365386  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OJIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
; CURRENT APPLICATION NUMBER: US/09/518,386B  
; CURRENT FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: EP 99104568.1  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: EP 00101666.6  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: TRANSIT  
; LOCATION: (1)..(26)  
US-09-518-386B-1

Query Match 100.0%; Score 2852; DB 3; Length 557;  
Best Local Similarity 100.0%; Pred. No. 7.1e-278;  
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MFILVLLTGALGLAFAFSWASIAFSLYLAPRRSSLYNLOGNHTNYFTGTFDILSARTG	60
Db	1	MFILVLLTGALGLAFAFSWASIAFSLYLAPRRSSLYNLOGNHTNYFTGTFDILSARTG	60
QY	61	EEHAKYREKYGSTLRFGAGAPVLNSTDPKVFNVHVKAEYDYPKPGMAARVARIATGDG	120
Db	61	EEHAKYREKYGSTLRFGAGAPVLNSTDPKVFNVHVKAEYDYPKPGMAARVARIATGDG	120
QY	121	VVTAGEAHKRRHIMIPSLSAQAVKSNVPITFLEKGMELVDKMMEDAAEKDMAYGESAGE	180
Db	121	VVTAGEAHKRRHIMIPSLSAQAVKSNVPITFLEKGMELVDKMMEDAAEKDMAYGESAGE	180
QY	181	KKATRLTEGVVDKDWGRATLDYMALAGDFYKSDSLQNKTNELYVAVGLTDCGFATLD	240
Db	181	KKATRLTEGVVDKDWGRATLDYMALAGDFYKSDSLQNKTNELYVAVGLTDCGFATLD	240
QY	241	SFKAIMDVFVYPTMKRRHEIPTQGLAVSRVRVGIELMEQKQAVLGASDAQVADKKDV	300
Db	241	SFKAIMDVFVYPTMKRRHEIPTQGLAVSRVRVGIELMEQKQAVLGASDAQVADKKDV	300

QY 301 QGRDILSLVVRANIANLPESOKLSDEEVLQAIISNLLFAGYETSSTVLTWFMHRLSEDKA 360  
 Db 301 QGRDILSLVVRANIANLPESOKLSDEEVLQAIISNLLFAGYETSSTVLTWFMHRLSEDKA 360  
 QY 361 VQKLRREICQIDTMDPTLDELNALPYLEAFVKSRLDPPSPYANRECLKDEDFIPLAE 420  
 Db 361 VQKLRREICQIDTMDPTLDELNALPYLEAFVKSRLDPPSPYANRECLKDEDFIPLAE 420  
 QY 421 PVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWLEDTVDSLSNIEA 480  
 Db 421 PVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWLEDTVDSLSNIEA 480  
 QY 481 PYGHOASFISGPRACFGWRFAVAKMAKFLVTLRRVQFEPILSHPEYEHITLIISPRIV 540  
 Db 481 PYGHOASFISGPRACFGWRFAVAKMAKFLVTLRRVQFEPILSHPEYEHITLIISPRIV 540  
 QY 541 GREKEGYQMRLOVKPVE 557  
 Db 541 GREKEGYQMRLOVKPVE 557  
 RESULT 2  
 US-09-518-386B-3  
 ; Sequence 3, Application US/09518386B  
 ; Patent No. 6365386  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HOSHINO, Tatsuo  
 ; APPLICANT: OJIMA, Kazuyuki  
 ; APPLICANT: SETOGUCHI, Yutaka  
 ; TITLE OF INVENTION: ASTAXANTHIN SYNTHETASE  
 ; FILE REFERENCE: ASTAXANTHIN SYNTHETASE  
 ; CURRENT APPLICATION NUMBER: US/09/518,386B  
 ; CURRENT FILING DATE: 2000-03-03  
 ; PRIOR APPLICATION NUMBER: EP 99104668.1  
 ; PRIOR FILING DATE: 1999-03-09  
 ; PRIOR APPLICATION NUMBER: EP 00101666.6  
 ; PRIOR FILING DATE: 2000-02-01  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; TYPE: PRT  
 ; LENGTH: 557  
 ; ORGANISM: Phaffia rhodozyma  
 US-09-518-386B-3

Query Match 100.0%; Score 2852; DB 3; Length 557;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-278; Mismatches 0; Indels 0; Gaps 0;  
 Matches 557; Conservative 0;  
 QY 1 MPILVLLTGALGALAFSASIAFTSLYLAPRRSSLYNLOGPNHTNYTGTGNFLDILSARTG 60  
 Db 1 MPILVLLTGALGALAFSASIAFTSLYLAPRRSSLYNLOGPNHTNYTGTGNFLDILSARTG 60  
 QY 61 BEHAKYREKYSTLRPAGIAGAPVLSNTPDKVFNHVMKEAYDYPKPGVAARVLRITAGDG 120  
 Db 61 BEHAKYREKYSTLRPAGIAGAPVLSNTPDKVFNHVMKEAYDYPKPGVAARVLRITAGDG 120  
 QY 121 VVTAAGEAHKRRRIMIPSLSAQAVKSMVPIFLEKGMELVDKRMEDAAEKDMAYGESAGE 180  
 Db 121 VVTAAGEAHKRRRIMIPSLSAQAVKSMVPIFLEKGMELVDKRMEDAAEKDMAYGESAGE 180  
 QY 181 KATRLTEGVVDKVGGRATLDVNALAGFDYKSDSLQNKTNELNLYVAVGLTDFGAPFLD 240  
 Db 181 KATRLTEGVVDKVGGRATLDVNALAGFDYKSDSLQNKTNELNLYVAVGLTDFGAPFLD 240  
 QY 241 SFKALMWDPVFPVFRMKRHEIPLTQGLAVSRVRGIELMEQKQAVLGSDQAVDKDV 300  
 Db 241 SFKALMWDPVFPVFRMKRHEIPLTQGLAVSRVRGIELMEQKQAVLGSDQAVDKDV 300  
 QY 301 QGRDILSLVVRANIANLPESOKLSDEEVLQAIISNLLFAGYETSSTVLTWFMHRLSEDKA 360  
 Db 301 QGRDILSLVVRANIANLPESOKLSDEEVLQAIISNLLFAGYETSSTVLTWFMHRLSEDKA 360

QY 361 VQKLRREICQIDTMDPTLDELNALPYLEAFVKSRLDPPSPYANRECLKDEDFIPLAE 420  
 Db 361 VQKLRREICQIDTMDPTLDELNALPYLEAFVKSRLDPPSPYANRECLKDEDFIPLAE 420  
 QY 421 PVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWLEDTVDSLSNIEA 480  
 Db 421 PVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWLEDTVDSLSNIEA 480  
 QY 481 PYGHOASFISGPRACFGWRFAVAKMAKFLVTLRRVQFEPILSHPEYEHITLIISPRIV 540  
 Db 481 PYGHOASFISGPRACFGWRFAVAKMAKFLVTLRRVQFEPILSHPEYEHITLIISPRIV 540  
 QY 541 GREKEGYQMRLOVKPVE 557  
 Db 541 GREKEGYQMRLOVKPVE 557  
 RESULT 3  
 US-09-144-367-2  
 ; Sequence 2, Application US/09144367  
 ; Patent No. 6432639  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lichter, Jay  
 ; APPLICANT: Guido, Marco  
 ; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
 ; FILE REFERENCE: SEQ-12P  
 ; CURRENT APPLICATION NUMBER: US/09/144,367  
 ; CURRENT FILING DATE: 1998-08-31  
 ; PRIOR APPLICATION NUMBER: 60/058,612  
 ; PRIOR FILING DATE: 1997-09-10  
 ; NUMBER OF SEQ ID NOS: 58  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 503  
 ; TYPE: PRT  
 ; ORGANISM: H. sapiens  
 US-09-144-367-2

Query Match 13.9%; Score 396.5; DB 4; Length 503;  
 Best Local Similarity 28.7%; Pred. No. 7.6e-31;  
 Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;  
 QY 13 LAAPSW--ASTAFPSLYLAPRS--SLYNLQ--PNHTNY--FTGNFLDILSARTG-----E 61  
 Db 7 LAWETWLLAVSLVLLTYGTHSHGLFKLGIPTPLFLGN---ILSYHKGCFMDE 63  
 QY 62 BEHAKYREKYSTLRPAGIAGAPVLSNTPDKVFNHVM--KEAYD---PKPGVAARVLRITAT 117  
 Db 64 CHKKYKVGWGF--YDG--QQPVLAITDPMIKTVLKECYSVFTNRPFGVGFMSKA- 117  
 QY 118 GGVVTAAGEAHKRRRIMIPSLSAQAVKSMVPIFLEKGMELVDKRMEDAAEKDMAYGES 177  
 Db 118 ---ISIADEEWEKRLRSLSPFTTSGKLEWVPIITAQIDVVLVNLREA----- 164  
 QY 178 AGEKATRLTET--EGVDVKDWVGRATLDVNALAGFDYKSDSLQNKTNELNLYVAVGLTDFGA 236  
 Db 165 -----ETCKPVLKDVFGAYSMDVITSTSGFNWIDSLNPNQD----- 201  
 QY 237 PTLDSFKAM--WDFV--PFRFMK--RRHEIPLTQGLAV---SRVGIELMEQKQAVLGSA 290  
 Db 202 PFVENTKLLRDFLDPPFLSITVPFFLIPILEVINICVFFREVNTFRKSVK- 256  
 QY 291 SQAQVXKDVQGRDILSLVVRANIANLPESOKLSDEEVLQAIISNLLFAGYETSSTVLTW 350  
 Db 257 KESRLDTQKRVDFLQIMDSQNSKETESHKALSDELVAQSIIIFIAGYETTSVLSP 316  
 QY 351 MFRHLSKXAVQKLRREICQI--DTDPTLDELNALPYLEAFVKSRLDPPSPYANRE 408  
 Db 317 IMVELATHPDVQKLRREICQI--DTDPTLDELNALPYLEAFVKSRLDPPSPYANRE 408  
 QY 409 CLKDEDFIPLAEFVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWML 468  
 Db 409 CLKDEDFIPLAEFVIGRDSVINEVRITKGTWVMLPLFNINRSKFIYGEDAEFPERWML 468

Db 377 CKKDE-----INGWIFPKGWWVMPISYALHRDP-KYWTPEKFLPERFS 420  
QY 469 EDVDTSLNS-TEAPYGHQASISPRACGWRFAVAEMKAFLEFVTLRRVQFEP 520  
Db 421 KKKONIDPIYTFPG-----SGPRNCIGWRFALMMKALIRVLQNFSPFKP 467

RESULT 4  
US-08-457-274A-25  
; Sequence 25, Application US/08457274A  
; Patent No. 5734086  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Jeffrey G.  
; APPLICANT: Iomica, Takashi  
; TITLE OF INVENTION: Cytochrome P450lpr Gene and Its Uses  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,274A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/240 (D-1519)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1304  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 504 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Rat  
; STRAIN:  
; DEVELOPMENTAL STAGE: Adult  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT:  
US-08-457-274A-25

Query Match 13.0%; Score 370; DB 1; Length 504;  
Best Local Similarity 26.2%; Pred. No. 3.5e-28;  
Matches 147; Conservative 84; Mismatches 189; Indels 140; Gaps 23;

QY 7 LTGALGLAFAFWASTAFSLY-LAPRRSSLYNLOQ-----PNTNYFTGNF-LDI 54  
Db 3 LLSALTLETWVLLAVLLVLYGFGTRTHGLFKKQGPGLPFPFGTVLYNYGLMKFDV 62  
QY 55 LSARTGEHAKYREKYGSTLRFAGIAGAPVLNSTDPKVFHVM-KEAY-----DYPKP 106  
Db 63 -----ECHKKYKLGW---LFDG--QMPLEAIDTETMIKNVLYKECFSVFTNRDFGPV 111  
QY 107 GVAARVLRATGDGVVTAEGEAKHRRIMIPSLSAQAVKSMWPIFLEKGNELVDKMMED 166  
Db 112 GI-----MGKAVSVARDEEWKEYRALLFTFTSGRLKEMFFIIEQYGDILVKYLQEE 163  
QY 167 AAEDKMAVCESAGEKATRLLET-EGVDVKDWVGRATLDVMALGFDYKSDSLQN----- 219

Db 164 A-----ETGKPVMTKKVFCAYSDVITSTSGVNVNDSLNNPKDPFV 204  
QY 220 -KTNEL-----YVAFVGLTDGPAPTLDSEFKAIW--DFVPYFRTMKRRHEIDL 264  
Db 205 EKYKLLRFFDFDLFLSVVLPFLT-----PIYEMLNICMPFKXDSIEFK----- 250  
QY 265 TQGLAVSRVRGIELMEQKQAVLGSASDQAVDKKDVQGRDILSLIVRANIAANLPSQ-K 323  
Db 251 -----KFVYRMKETRLDSVOKHRV-----DFLOLMNVAHNDKSKESHTA 290  
QY 324 LSDEVLQAISNLLFAGYETSSVLTWMEHRLSEDAVODKLEEEICQI--DTDMPDLDE 381  
Db 291 LSDNEITAQSIIFIFAGYEFTSSTLSFVLSLATHEDTKKLEEDRALPNKAPFYDT 350  
QY 382 LNALPYLEAFVKESLRDPPSPYANRECLKDEDFIPLABPVGDRDGSVINEVRIKGTWV 441  
Db 351 VMEMEYLDMLNETLRLYPIGNRLERYCKKDE-----INGVFMPPKGSVV 395  
QY 442 MLPLFNINRSKFIYGEDAEERFERMLEDVTDLSNS-LEAPYGHQASFTSGPRACGWRP 500  
Db 396 MIPSYALHRDPOHWPE-PEEFRPERFSKENKGSIDPYVLYLPGF-----NGPRNCIGMEF 448  
QY 501 AVAEMKAFLEFVTLRRVQFEP 520  
Db 449 ALMMNKLALTKVLQNFSPFQ 468

RESULT 5  
PCT-US95-05758-25  
; Sequence 25, Application PC/TUS95050758  
; GENERAL INFORMATION:  
; APPLICANT: Cornell Research Foundation, Inc.  
; TITLE OF INVENTION: Cytochrome P450lpr Gene and Its  
; TITLE OF INVENTION: Uses  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/05758  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1304  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 504 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Rat  
; STRAIN: Unknown  
; DEVELOPMENTAL STAGE: Adult  
; POSITION IN GENOME:

CHROMOSOME/SEGMENT: Unknown  
PCT-US95-05758-25

Query Match 13.0%; Score 370; DB 5; Length 504;  
Best Local Similarity 26.2%; Pred. No. 3.5e-28;  
Matches 147; Conservative 84; Mismatches 189; Indels 140; Gaps 23;

QY 7 LTCALGLAFAWSIAFTSLY-LAPRRSLYNLQ-----PNHTNYFTGNF-LDI 54  
DB 3 LLSALTLETWLLAVLVLLYFGFRTGLFKKQIGPCPKLPFFGVVLYVYGLKFEV 62  
QY 55 LSARTGEHAKREKYGSTLRAGTAGAPVUNSTDPKVFNVHM-KEAY-----DYPKP 106  
DB 63 -----ECHKYKGIWNG---LFDG--QMLPFAITDTEMIKIVLKECFVSFTNRDPGV 111  
QY 107 GMAARVLRIATGDGVVTAEGEAKHRRIMIPSLSAQAQVKSMPVIFLEKGMELVDKMD 166  
DB 112 GI-----MGKAVSVAKEEMKRYRALLSPTTSGFLKEMFPII BOYGDILVKYLQE 163  
QY 167 AAEDKAVGESAGEKATRLT-EGVYVYKDVWVGRATLDVVALAGFDYKSDSLQ----- 219  
DB 164 A-----ETGKEVTWKVYFGAYSMDVITSTSGVNVDSLPNNPKDPFV 204  
QY 220 -KTNEL-----YVAFVGLTDGFAPTLDSFRAIMW--DFVPYFRTMKRHEIPL 264  
DB 205 ETKTKLLRDFDPLFLSVLFPPLT-----ETGKEVTWKVYFGAYSMDVITSTSGVNVDSLPNNPKDPFV 264  
QY 265 TOGLAVSRVGTLEMEQKQAVLGSDASQAVDKDVQGRDILSLVLRANLANLPESQ-K 323  
DB 251 -----KFYRKETRLDSVQHRV-----DFLQMMNAHNSDKESHTA 290  
QY 324 LSDEEVLQISNLLFAGYETSSTVLTMMFHLRSEDKAVQDKLREEICQI--DTDMPTLDE 381  
DB 291 LSDMEITAGSIIFIFAGYEPTSTLSFVLHSLATHPTQKQLQEBIDRALPNKAPPYDT 350  
QY 382 LNALPYLEAFVKSRLDPPSPYANRECLDEDFIPLAEFVIGRGGVINEVTRITKTMV 441  
DB 351 VMEMEYLDVNLNETLRLYPIGNRERVCKDVE-----INGVFMKGSVV 395  
QY 442 MLPFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEAPYHQASFIIS-IEAPYHQASFIISGPRACFWRP 500  
DB 396 MIPSVLALHRDPQHWPE--PEERFPERFSKENKSGSIDPYVLPFG-----NGPRNCIGMRP 448  
QY 501 AVAEMKAFILVTLRRVQFEP 520  
DB 449 ALMMNKALTKVLQNFSPQ 468

RESULT 6  
US-09-527-073-2  
; Sequence 2, Application US/09527073  
; Patent No. 653433  
; GENERAL INFORMATION:  
; APPLICANT: Michael M. Neff  
; APPLICANT: Joanne Chory  
; TITLE OF INVENTION: GENETICALLY MODIFIED PLANTS HAVING  
; TITLE OF INVENTION: MODULATED BRASSINOSTEROID SIGNALING  
; FILE REFERENCE: SALKINS 024A  
; CURRENT APPLICATION NUMBER: US/09/527,073  
; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: US 60/124570  
; PRIOR FILING DATE: 1999-03-16  
; PRIOR APPLICATION NUMBER: US 60/170,931  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: US 60/172,832  
; PRIOR FILING DATE: 1999-12-20  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-527-073-2

Query Match 12.9%; Score 367; DB 4; Length 520;  
Best Local Similarity 22.2%; Pred. No. 7.5e-28;  
Matches 128; Conservative 111; Mismatches 203; Indels 134; Gaps 21;

QY 4 LVLITGALGLAFAWSIAFTSLY-LAPRR-----SSLYNLOGFNHTNYFTGNFELDILS--- 56  
DB 13 VLVLSSVLSLVIVKMSL---LWVRPKIEEHFSGQIRGPPY-HFFIGNVKELVGMWL 67  
QY 57 -----ARTGESHAKEKYGSTLRAGTAGAPVUNST--DPKVFNVHMKEA 100  
DB 68 KASHPMFPFNNILPVLVSFYHWRKIYGATFL---VWFGPTFLTVADPDLIREIFS 124  
QY 101 YDYPKPGMAARVLRIATGDGVVTAEGEAKHRRIMIPSLSAQAQVKSMPVIFLEKGMELV 160  
DB 125 EHYEK-NEAHLPLVKQEGDGLLSLKGSKWAHREKIIISPTFHMENLKLVPVVLKSVTDMV 183  
QY 161 DKMMEDAAEKDMVAGESAGEKATRLTEGVYVYKDVWVGRATLDVVALAGFDYKSDSLQNK 220  
DB 184 DKMSDKLSEN-----GEVE-----VDYVWFQILTEDVISRTAFSGSYE----- 222  
QY 221 TNELYVAFVGLTDGFAP-----TLDSEKALIMWDFVYFRTMKRHEIPLTQGLAV 270  
DB 223 -----DGRAVRLQAQOMLLCBAFAFOV---FIPGYRFPF-----TRGNLK 260  
QY 271 SRRVG-----IEMEQKQAVLGSDASQAVDKDVQGRDILSLVLRANLANLPESQK 323  
DB 261 SRKLDKIRKSLKLLKIRRRQNAIDGEGEC---KEPAKDLGLMLQA-----KN 308  
QY 324 LSDEEVLQISNLLFAGYETSSTVLTMMFHLRSEDKAVQDKLREEICQI--DTDMPTLDE 381  
DB 309 VTVDIVVECKSPFPFAGKQTTSNLLTWTLLSNHPEWQAKARDEVLRVCGSRDVPKDH 368  
QY 382 LNALPYLEAFVKSRLDPPSPYANRECLDEDFIPLAEFVIGRGGVINEVTRITKTMV 441  
DB 369 VVKLTLSMINESLRLYPVIVATIRAKSDVK-----LGGYKIPCGTEL 413  
QY 442 MLPFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEAPYHQASFIIS---GPRACFW 498  
DB 414 LIPTIAVHHDQAIWGNVDVNEFPARFADGVPRAAK-----HPVGFIPFGLGVRTCIG 466  
QY 499 RFAEMKAFILVTLRRVQFEPFIISHPEYHITLI 534  
DB 467 NLAILQKLLVLMVIOQRTFH---LAPTYQAPTVL 499

RESULT 7  
US-09-976-594-533  
; Sequence 533, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 533  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 6825202CD1  
US-09-976-594-533

Query Match 12.1%; Score 345; DB 4; Length 524;  
Best Local Similarity 23.4%; Pred. No. 1.2e-25;  
Matches 140; Conservative 104; Mismatches 207; Indels 146; Gaps 25;



```

172 MAVGESAGEKKATRIULETGVGVKVMVGRATLDVNMALAGFDYKSDSLQNKTNELNYVAFVGL 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 F-----KHVRKHGQTFDIQELFPLRTVDSATEBFLGESAESLRDSD-----VGL 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 TUGFAPTLDSFKAIKWDVPVPTWK--RRHEIPLTQGLAVSRVRGIELMELMEQKQAVLGS 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 T-----PTTKDFEG-RGDFADAFNYSQTYQAYRFLQOMYILN--GAEF--RKSIAIVHK 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
290 ASDQAVDK-----XDVGRDILSLLVRAANIAANLPSQKLSDBEVLAA-QISNLLFAGYE 342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
288 FADHYVQKALELTDLDLQKQGVFLYEL-----AKQTRDPKVLRDQLNLVAGRD 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
343 TGSTVLTMFHLRSLSDKAVQDKLREEI-----CQIDTMDPTLDELNALPYLEAFVKGS 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 TTAGLLSFVFEYELSRNPVFAKLRESEVENRFGLGSEARVBEISFESLKSCEYLKAVINEA 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 LRLDPPSPANRECLKDDEFPLAEPVLGRGGSVINEVRITKGMVMWLPLFNINRSKFY 455
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 LRLYPSVPHNFVATNTT-LFRGG-----GKDG--CSPIVVKKGQVMYTWIGTHRDPDIY 453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 GEDABEFRERMLWDVTDLSLSIEAPYGHQASFGSPRACFGMRFAVAEMKAFLFVTLRR 515
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 GADADVFRERFERFEETRKLGHAYVP-----FNGGPRCLGQGFALTEAS-----YVTVRL 504
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
516 VQ-----PEPIISHPEYEHITLIIS 535
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 LQFEGNLSLDPNAEYPPKPLONTLTLS 530
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US 08-457-274A-22
US 08-457-274A-22
; SEQUENCE 22 Application US/08457274A
; PAST NO. 5733086
; GENERAL INFORMATION:
; APPLICANT: Scott, Jeffrey G.
; APPLICANT: Tomita, Takashi
; TITLE OF INVENTION: Cytochrome P4501b1 Gene and its Uses

```

[illegible]

```

RESULT 11
PCT-US95-05758-22
Sequence 22, Application PC/TUS9505758
GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: Cytochrome P450lpr Gene and Its
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: P. O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05758
FILING DATE:

```

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael J.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Musca domestica
; STRAIN: Rutgers
; DEVELOPMENTAL STAGE: Adult
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: Chromosome 5
PCT-US95-05758-22

Query Match 11.6%; Score 330.5; DB 5; Length 507;
Best Local Similarity 23.5%; Pred. No. 3.4e-24;
Matches 131; Conservative 94; Mismatches 210; Indels 123; Gaps 19;

Qy 10 ALGLAFAFWASIAEFSLYAPRRSSLYNLQPNH--TNYTGNFLDLSA-RTGEEHAKY 66
Db 10 ALGV-----LASLAFYVRWNGFYWKRRGIPHEEPHLVNGVNGKLSRYHIGETIADY 62

Qy 67 REKYGSTRFAGI--AGAPVLNSTDPKVFNVHM-KEAYDYPKPGMAARVLRIATGQVVT 123
Db 63 YRKFGSGPPAGIFLGHKPAAVLVDKELRKVLKDFSNFANRGLYNEKDDPLTGLVLM 122

Qy 124 AEGEAHRRHRIMIPSLSAQAVKSMVPIFEKGMELVDKMDAAEKDVAAGSAGEKKA 183
Db 123 VEGEKWRLTKLSPTTAGCKYMYNTVLEVGQRLLEVME----- 164

Qy 184 TRLETEG-VVDKQVGRATLDVMAAGFDYKSDSLQK-----KTNELYV 226
Db 165 -KLEVSSELDWRDLARFNTDVGSAFGTECSLRNPHRFLAMGRKSTIEVPRHNLIM 223

Qy 227 AFVLGTGCAPTLDSFKAKMWDVFVYFRTWKRRHEIPLTQGLAVSRVVG-ELMEQKQAV 286
Db 224 AFI-----DSFPPEL-----SRKLGMRVLPEVDVHQF 248

Qy 287 LGSASDOAVD---KKDVQGRDILSLVVRANIAANLPESQK-----LSDVEVLAQISNLLFA 339
Db 249 FMSIKETVYREKNIRNRDFLDVLDEK---NNPESISKGLGLTFNELAAQVFFPLG 305

Qy 340 GYETSSTVLTWPHRLSEDKAVQDKLREEI---CQIDTDMPTLDLNLALPYLEAFVKES 395
Db 306 GFETSSSTMGFALYELAQNLQDRLEEVNEVDFQPKEDNISYDALMNIPIYLDQVNLNET 365

Qy 396 LRLDPPSPYANRECKDEDFIPLAEPVIGRDGSGVINEVITKTMTWMLPFLNINRSKFIY 455
Db 366 LRKYPVGSALTQTLNDY-----VVPHPKYV-----LPKGLTFVIFVLGIHYDPELY 413

Qy 456 GDAEEFRPRRWLEDDVDSLSNEAPYGHQASFIISGPRACFGWRFAVAEMKAFLEFVTLRR 515
Db 414 -PNPEEDPERSPEMVKQDSDV-----WJFGDGFRCNIGRFRGMSRSLGLAIVIRH 467

Qy 516 VQPE-----PIISHPE 526
Db 469 FRFTVCSRTDIPMQINPE 485

RESULT 12
US-09-583-447A-4
; Sequence 4, Application US/09583447A

; Patent No. 6645745
; GENERAL INFORMATION:
; APPLICANT: WOJNOWSKI, Leszek
; APPLICANT: GELLNER, Klaus
; APPLICANT: EISELT, Regina
; TITLE OF INVENTION: IDENTIFICATION OF A NEW MEMBER OF THE CYTOCHROME P450 3A
; TITLE OF INVENTION: (CYP3A) GENE FAMILY: CYP3AX
; FILE REFERENCES: 310115.401
; CURRENT APPLICATION NUMBER: US/09/583,447A
; CURRENT FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-583-447A-4

Query Match 11.5%; Score 328; DB 4; Length 504;
Best Local Similarity 27.1%; Pred. No. 6e-24;
Matches 127; Conservative 75; Mismatches 185; Indels 82; Gaps 18;

Qy 68 EKYGSTRFAGI--AGAPVLNSTDPKVFNVHM-KEAYDYPKPGMAARVLRIATGQVVT 124
Db 66 EKYG---EMWGLYEGQQPMLVIMDPDMIKTVLVKCYSVFTNQPLGPMGFLK-SALSFA 121

Qy 125 EGEAHEHRRIMIPSLSAQAVKSMVPIFEKGMELVDKMDAAEKDVAAGSAGEKKA 184
Db 122 EDEEWRKIRTLSPASTSVKFKEMVFIISQCGDMLVRSLOEAE----- 165

Qy 185 RLETEGVDVQVGRATLDVMAAGFDYKSDSLQKTNELYVAVGLTGDGFAPTLDSFKA 244
Db 166 --NSXSINLKDPFGAYTMDVTGTLFGVNLDSLNNPQD-----PFLKNMKK 209

Qy 245 IM-WFVPVYFRTWKRRHEIPLTQGLA-VSRVVGIELMEQKQAVLGSASDAQVDK--KDV 300
Db 210 LLKDLDFLPFLL-----ISLFPFLTPVFEALNIGLFPKDVTHFLKNSIERMKESRLKDK 264

Qy 301 QCR---DILSLVVRANIAANLPESQKLSDBEVLQISNLLFAGYETSSTVLTWPHRLSED 358
Db 265 QKRVDFVFFQOMIDSONSKETSKHLSLELVAQSIIIFAYDTTSTLPIFMYELATH 324

Qy 359 KAVQXKLREBEICQIDTMP-----TLDELNALPYLEAFVKBSLRLDPPSPYANRECKDE 413
Db 325 PDVQOKLOBE---IDAVLPNKAPVTDALVQMEYLDVMVNETLRLFPVSVRTRVCKKI 381

Qy 414 DIPLAEPVIGRDGSGVINEVITKTMTWMLPFLNINRSKFIYGDABEFRE-RLVEDVT 472
Db 382 E-----INGVFIPKGLAVVPIYALHHDHP-KYWTPEPEKFCPESRFSKQK 425

Qy 473 DLSNSIE-APYGHQASFIISGPRACFGWRFAVAEMKAFLEFVTLRRVQFEP 520
Db 426 DSDIDLYRYIPFG-----AGPRNCIGMFAFALTNIKLAIVIRALQNFSEKP 468

RESULT 13
US-09-302-620B-98
; Sequence 98, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eschco, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS

```







```
; FEATURE:
; OTHER INFORMATION: Altered sequences
US-09-158-767-20

Query Match      11.4%; Score 326.5; DB 3; Length 541;
Best Local Similarity 24.6%; Pred. No. 9.5e-24;
Matches 125; Conservative 87; Mismatches 197; Indels 99; Gaps 20;

Qy 78 GIA--GAPVLNSTDPKVFNVHVMKEAYD-YKPGMAARVLRATGDGVVTAEGEBAHKHRR 134
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 73 GVARRGGGLVTVTCDFRNLEHVLKARFDNYKPGPFHGVFRDLLGDGIFNSDGDITWLAQRK 132
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 135 IMIFSLSAQAVKSVPIFLKGMELVDKMDAEBKMAVGESAGEKKAIRLETEGVGVK 194
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 TAALEFTRTURTAMSVRSIIH--GRLLPILA--DAAKGKAQ-----VDLQ 176
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 195 DWGCRATLDVVALAGFDYKDSLQN--KTNELYVAFVGLTDGPAPTLDSFKAIMWDFVPY 252
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 177 DLLRLAFDNLICGLAFGKDPETLAQGLFENEFASAFDRATEA---TLNRF-----IFPE 227
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 253 F-----RTMKRRHEIPTQGLA-VSRVGIELMEQKQAVLGASDAQVKKDVOGGRDIL 306
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 228 FLWRCKKWLGLGEMTTLTSSMAHVDQYLAIVIKRKLKLEAAGNGKCDTAATHD----DLL 283
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 307 SLLVRANIAANLPESQKLSDEVLQISNLLFAGYETSSVLTWMPHRLSEDKAVQDKLR 366
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 284 SRFRKRG-----SYSDLSQHVAINFLAGRTSSVALSWFFWLVSTHPAVERKIV 334
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 367 EEICQI-----DTDMPITLDELNALPYLEAFVKESLRLDPPSPYANRECLKDE 413
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 335 RELCSVLAASRGADHPALWLAEPFTFEELDRLVYLKAALSETLRLYPSPVPEDSKHVVD- 393
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 414 DFIPLAEPVIGRDSVINEVRITYGTWMLPLFNINRSKFIYGEDAEFRPERWLE----- 469
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 394 DYLP-----DGTG-----VPAGSSVTYSIYAGRMKGWNGBDCLFEPERWLSADGT 440
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 470 --DVTDSINSIEAPYGHQASFGFRACFGWRFAEMKAFLFVTLRRVQFEPIISHPY 527
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 441 KFECHDSYKVV-----AFNAGPRVCLGKDLAYLQMKNIAGSVLLRHRLTVAPGRHVE 492
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 528 EHITLIISRPRIVGREKEGYQMRLOQVP 555
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 493 QKNSLTLM-----KGG--LRMEVRP 511
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: December 14, 2004, 21:55:12  
Job time : 40 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2004, 14:27:33 ; Search time 150 Seconds  
(without alignments)  
1332.081 Million cell updates/sec

Title: US-10-066-007a-1  
Perfect score: 2852  
Sequence: 1 MFILVLTGALGLAAFWAS.....RIVGREKEGYQMRLOKVPVE 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Genesec 23Sep04: \*  
1: Genesecp1980a: \*  
2: Genesecp1990a: \*  
3: Genesecp2000a: \*  
4: Genesecp2001a: \*  
5: Genesecp2002a: \*  
6: Genesecp2003a: \*  
7: Genesecp2003bs: \*  
8: Genesecp2004a: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2852	100.0	557	3	AAB08713 Amino aci
2	414	14.5	503	8	ADO47245 Pig cytoc
3	404	14.2	503	8	ADO47244 Dog cytoc
4	402	14.1	501	8	ADO47242 Rabbit cy
5	400	14.0	485	8	ADJ87521 Human cyt
6	399	14.0	503	8	ADO47238 Human cyt
7	398.5	14.0	503	6	ABU57260 Human chy
8	398.5	14.0	503	6	ABR82024 Human cyt
9	398.5	14.0	503	8	ADL18578 Human wil
10	398.5	14.0	503	8	ADL18672 Human cyt
11	398	13.9	503	8	ADO47246 Murine cy
12	396.5	13.9	502	7	ADD48378 Human Pro
13	396.5	13.9	503	2	AAR72363 Human cyt
14	396.5	13.9	503	2	AAR81464 Human der
15	396.5	13.9	503	2	AAR93170 Human cyt
16	396.5	13.9	503	2	AAV05302 Human CYP
17	394	13.8	485	8	ADO33768 Human CYP
18	394	13.8	503	5	ABG68753 Cytochrom
19	392.5	13.8	527	5	ABO8079 Maize cyt
20	390.5	13.7	502	8	ADO47239 Human cyt
21	385.5	13.5	503	5	AAE22852 Human cyt
22	385.5	13.5	503	8	ADO47243 Human cyt
23	385.5	13.5	535	5	ABG68747 Cytochrom
24	385	13.5	502	5	AAE26192 Human cyt
25	385	13.5	502	7	ADE57186 Human Pro

26	385	13.5	502	7	ADE57190 Human Pro
27	385	13.5	502	7	ADE53919 Human Pro
28	385	13.5	502	8	ADJ75595 Marker ga
29	385	13.5	502	8	ADN04176 Antipsori
30	385	13.5	502	8	ADO47240 Human cyt
31	384.5	13.5	513	8	ADN10687 Nicotiana
32	382	13.4	454	5	ABG68754 Cytochrom
33	374.5	13.1	461	8	ADN10591 Nicotiana
34	373.5	13.1	454	6	ADA48312 Rice prot
35	370	13.0	504	7	ADE57188 Rat Prote
36	370	13.0	504	7	ADE53917 Rat Prote
37	370	13.0	504	7	ADE57184 Rat Prote
38	370	13.0	504	8	ADO47237 Rat cytoc
39	370	13.0	521	8	ADN10631 Nicotiana
40	367	12.9	520	3	AAB23917 Arabidops
41	367	12.9	521	8	ADN10679 Nicotiana
42	365	12.8	520	6	ABU54571 Human NOV
43	363.5	12.7	512	5	AAU97096 Abecisic
44	362.5	12.7	465	3	AAAG54205 Arabidops
45	362.5	12.7	475	3	AAAG54204 Arabidops

## ALIGNMENTS

RESULT 1  
AAB08713  
ID AAB08713 standard; protein; 557 AA.

XX  
AC AAB08713;  
XX  
DT 12-SEP-2003 (revised)  
DT 02-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of an astaxanthin synthetase polypeptide.  
XX  
KW Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;  
KW antioxidant; cancer; colouring reagent; farmed fish; salmon.  
XX  
OS Xanthophyllomyces dendrorhous.  
XX  
PN EP1035206-A1.  
XX  
PD 13-SEP-2000.  
XX  
PF 03-MAR-2000; 2000EP-00104430.  
XX  
PR 09-MAR-1999; 99EP-00104668.  
XX  
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
PI Hoshino T, Ojima K, Setoguchi Y;  
XX  
DR WPI; 2000-559874/52.  
XX  
DR N-PSDB; AAA64472, AAA64473.  
XX  
PT Novel polynucleotide encoding astaxanthin synthase useful for producing  
XX recombinant cells for producing astaxanthin from beta-carotene.  
XX  
PS Claim 2; Page 21-23; 46pp; English.

applicable

The present sequence represents an astaxanthin synthetase polypeptide of Phaffia rhodzyma. The enzyme is involved in the last step of the astaxanthin biosynthesis pathway from beta-carotene to astaxanthin. P. rhodzyma is a carotenogenic yeast strain. The astaxanthin synthetase polynucleotides and polypeptides are useful for producing astaxanthin. CC Astaxanthin is an antioxidant which may be used to protect living cells against diseases such as cancer. Astaxanthin is also used as a colouring reagent, e.g. in farmed fish like salmon to impart an orange-red coloration. (Updated on 12-SEP-2003 to standardise OS field)

Sequence 557 AA;

Query Match	100.0%;	Score 2852;	DB 3;	Length 557;
Best Local Similarity	100.0%;	Pred. No. 4.8e-240;		
Matches 557;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTLVLLTALGALAFSPWASIAFFSLYLAPRSSLYLNQGNHNTYFTGNFLDLSARTG	60	
Db	1	MTLVLLTALGALAFSPWASIAFFSLYLAPRSSLYLNQGNHNTYFTGNFLDLSARTG	60	
QY	61	EEHAKREKYGSTLRFPAGI--ACAPVLNSTDPKVFNVHVMKEAYDYDPKPGMAARVLRIATGDG	120	
Db	61	EEHAKREKYGSTLRFPAGI--ACAPVLNSTDPKVFNVHVMKEAYDYDPKPGMAARVLRIATGDG	120	
QY	121	VVTAEGEAHRRHRIIMPISLSAQAVKSWPIFLEKGMELVDKMMEDAAEKDMAGVSGAGE	180	
Db	121	VVTAEGEAHRRHRIIMPISLSAQAVKSWPIFLEKGMELVDKMMEDAAEKDMAGVSGAGE	180	
QY	181	KKATRLTEGVVDKVGWRATLDVVALAGFDYKSDSLQNKTNELYAFVGLTDFGAPTLD	240	
Db	181	KKATRLTEGVVDKVGWRATLDVVALAGFDYKSDSLQNKTNELYAFVGLTDFGAPTLD	240	
QY	241	SPKAIMWDVPVFRFMKREHEIPLTQGLAVSRVRYGIELMEOKKCAVLGSASDAQVKKOV	300	
Db	241	SPKAIMWDVPVFRFMKREHEIPLTQGLAVSRVRYGIELMEOKKCAVLGSASDAQVKKOV	300	
QY	301	QGRDILSLVRANIAANLPESQKLSDEEVLQAQISNLLFAGYETSTVLTWTFHRLSEDKA	360	
Db	301	QGRDILSLVRANIAANLPESQKLSDEEVLQAQISNLLFAGYETSTVLTWTFHRLSEDKA	360	
QY	361	VODKREECOTDPTDELNALPYLEAFVKSRLDPSVANECLKDEDFPLAE	420	
Db	361	VODKREECOTDPTDELNALPYLEAFVKSRLDPSVANECLKDEDFPLAE	420	
QY	421	PVIGRDGVSINVRITKGTMMVLPFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA	480	
Db	421	PVIGRDGVSINVRITKGTMMVLPFNINRSKFIYGEDAEFRPERWLEDVTDLSNIEA	480	
QY	481	PYGHQASFTSGPRACFGRRPAVAEMKAFVTLTARVQPEPIISHPEYEHITLIISPRIV	540	
Db	481	PYGHQASFTSGPRACFGRRPAVAEMKAFVTLTARVQPEPIISHPEYEHITLIISPRIV	540	
QY	541	GREKEGYQMLQKXPE 557		
Db	541	GREKEGYQMLQKXPE 557		
RESULT 2				
ID	AD047245	AD047245 standard; protein; 503 AA.		
AC	AD047245;			
XX	29-JUL-2004	(first entry)		
DE	Pig cytochrome P450 CYP3A29	SeqId 17.		
XX	dynamics docking; cytochrome P450; multispecific enzyme;			
KW	protein coordinate data; common structural block; CSB; pig.			
XX	Sus sp.			
XX	WO2004038655-A2.			
XX	06-MAY-2004.			
XX	28-OCT-2003; 2003WO-IB005134.			
XX	28-OCT-2002; 2002US-0421569P.			
XX	(COMS ) COMMISSARIAT ENERGIE ATOMIQUE.			
PA	(CNRS ) CNRS CENT NAT RECH SCI.			
XX	Andre F, Delaforge M, Loiseau N;			
PI				

XX	WPI; 2004-389974/36.			
DR	Designing 3 dimensional (3D) protein model, by identifying common			
XX	structural blocks (CSBs) among family members, aligning sequence,			
PT	defining 3D structure of CSBs, global constraints, selecting rotamers,			
PT	determining, optimizing 3D structures.			
XX	Claim 13; SEQ ID NO 17; 193pp; English.			
PS	This invention relates to a novel method for performing restrained			
XX	dynamics docking of one or more substrates on multi-specific enzymes.			
CC	Specifically, it refers to determining the three-dimensional structure of			
CC	active sites that are flexible and can adapt to different substrates i.e.			
CC	multispecific enzymes such as cytochrome P450. The present invention			
CC	describes identifying common structural blocks (CSBs) among members of			
CC	the family, aligning primary amino acid sequences, aligning the protein			
CC	as compared on the first alignment to obtain a second alignment, defining			
CC	3D structure of protein CSBs, defining global constraints and selecting			
CC	rotamers, in order to determine the family of 3D model protein structures			
CC	and optimise models by dihedral angles. Accordingly, the method can be			
CC	used for screening, designing or identifying natural, unnatural or			
CC	substrate analogues, as well as inhibitors, activators or modulators of			
CC	the multispecific enzyme in question. In addition, it can determine the			
CC	effect of a first substrate on a second substrate, which can then be			
CC	applied to pharmaceutical products. Furthermore, the method can be used			
CC	to determine the oxidative modification of the substrate according to its			
CC	proximity to a haem molecule, for performing dynamic docking of the			
CC	metabolite either in the absence or presence of a second substrate in the			
CC	computed simulation and to compare the energy of the bound metabolite			
CC	relative to the energy of its parent substrate bound, in order to			
CC	determine if the exit of the given metabolite from the enzyme is favoured			
CC	or not. This polypeptide sequence is a protein from the cytochrome P450			
CC	family of enzymes, given in an exemplification of the invention.			
XX	Sequence 503 AA;			
QY	Query Match	14.5%;	Score 414;	DB 8; Length 503;
Db	Best Local Similarity	27.4%;	Pred. No. 4.7e-27;	
QY	Matches 158;	Conservative 85;	Mismatches 197;	Indels 136; Gaps 23;
QY	12	GLAAFWASIA--FFSLYLAPRS-----SLYNLQGNHNTYFTGNFLDLSARTGEEH--	63	
Db	6	GFSTETWLLATSLVLLYLYGTYSGLFKKLGIPGPRPLPYF-GN---ILGVRKGVDPD	61	
QY	64	AKYREKYGSTLRFPAGI--ACAPVLNSTDPKVFNVHVM-KEAY-----DYPKPGMAARVL	113	
Db	62	KKCFQYQ--KMGVTDGRQPLLATDPMNKSVLVKCYSVFTNRRSGFLGAMRNAL	118	
QY	114	RIATGDGVVTAEGEAHRRHRIIMPISLSAQAVKSWPIFLEKGMELVDKMMEDAAEKDWA	173	
Db	119	SL-----AEDEWKRIITLLSPTFTSGKLKEMFPIISHYVDLLVSNLRKE-AEKG--	167	
QY	174	VGESAGEKATRLTEGVVDKVGWRATLDVVALAGFDYKSDSLQNK-----TNELYAF	228	
Db	168	-----KFTWKIDIFGAYSDVITSTAFGVNIDSLNPNQDPFVENSCKLLK	212	
QY	229	VGLTGDGAPTLDGFKAIMWDF-----VPYF-RTMKRHEIPIITQGLAVSRVVG	275	
Db	213	FSFDPPELLSLIIFPFLTPIFEVINITLFPKSVNFFTSKVKMKESRLT-----	262	
QY	276	IELMEQKQAVLGASDAQVDDKQGRDILSLVRANIAANLPESQKLSDEEVLQAQISN	335	
Db	263	-----DQQRV-----DLQMINSONSKEMDPKHSLSNBEELVAQGI	301	
QY	336	LLFAGYETSTVLTWTFHRLSEDKAVQDKLREICQI--DTMPTLDELNALPYLEAFVK	393	
Db	302	FIFAGYETTSALSLLAYELATHPDVQKQLEBIEATFFNKAPPTYDALAQEYELDMVN	361	
QY	394	BSLRLDPPSPVANECLKDEDFPLAEFVIGRDGVSINVRITKGTMMVLPFNINRSKF	453	
Db	362	ETRLYPIARLERACKDVE-----IHGVFVPKGVVVVVFVLRDPP	406	

[illegible]

RESULT 3  
ADO47244  
ID ADO47244 standard; protein; 503 AA.

AC ADC47244;

XX  
DT 29-JUL-2004 (first entry)

XX  
DE Dog cytochrome P450 CYP3A12 SeqID 16.

XX dynamics docking; cytochrome P450; multispecific enzyme;  
KW protein coordinate data; common structural block; CSB; dog.

Canis sp.

XX PN WO2004038655-A2.

XX  
PD 06-MAY-2004.

XX  
28-OCT-2003:

XX  
PR 28-OCT-2002: 2002IIS-0421569D

XX  
PA (COMS ) COMMISSARIAT ENERGIE  
XX  
LE 1007, RUE DE LA LIBERTE

PA (CNRS) CNRS CENT NAT RECH SCI.

PI Andre F, Delaforge M, Loiseau

DR WPI; 2004-389974/36.

xx. Designing 3 dimensional (3D) protein model, by identifying common structural blocks (CSBs) among family members, aligning sequence, defining 3D structure of CSBs, global constraints, selecting rotamers, determining, optimizing 3D structures.

XX  
PS Claim 13; SEQ ID NO 16; 193pp; English.

This invention relates to a novel method for performing restrained dynamics docking of one or more substrates on multi-specific enzymes. Specifically, it refers to determining the three-dimensional structure of active sites that are flexible and can adapt to different substrates i.e. multispecific enzymes such as cytochrome P450. The present invention describes identifying common structural blocks (CSBs) among members of the family, aligning primary amino acid sequences, aligning the protein as compared on the first alignment to obtain a second alignment, defining 3D structure of protein CSBs, defining global constraints and selecting rotamers, in order to determine the family of 3D model protein structures and optimise models by dihedral angles. Accordingly, the method can be used for screening, designing or identifying natural, unnatural or substrate analogues, as well as inhibitors, activators or modulators of the multispecific enzyme in question. In addition, it can determine the effect of a first substrate on a second substrate, which can then be applied to pharmaceutical products. Furthermore, the method can be used to determine the oxidative modification of the substrate according to its proximity to a haem molecule, for performing dynamic docking of the metabolite either in the absence or presence of a second substrate in the computed simulation and to compare the energy of the bound metabolite relative to the energy of its parent substrate bound, in order to determine if the exit of the given metabolite from the enzyme is favoured or not. This polypeptide sequence is a protein from the cytochrome P450 family of enzymes, given in an exemplification of the invention.

Sequence 503 AA;

Designing 3 dimensional (3D) protein model, by identifying common structural blocks (CSBs) among family members, aligning sequence,

				Query Match	14.2%; Score 404; DB 8; Length 503;	
				Best Local Similarity	29.0%; Pred.No 3.5e-26;	
				Matches 146; Conservative	83; Mismatches 184; Indels 90; Gaps 20;	
QY	69	KYGSTLPAGTAGAPVNSTDPKVFNVHM-KEAYDYPKPGBMAARVLRIATG-----DGVT	121			
Dd	67	KYGRNWGYD--GROFVLAITDDPDIKTVLKECY-----SVFTNRRTLLGPVGFMKSAI	118			
QY	122	VTAEGEAHKRRHRIWIFSLSAQAQVKSMVIPLEKGMELVDMQMDAAEKDKMVAVESAGEK	181			
Dd	119	SLSDEEWKRRTILLSPTFTTGLKENFFPIIGQGVDVLNNLRKE-AEKGKA-----	169			
QY	182	KATRLTEGVADVKNWGRAILDNMALAGFYKDSLOKNTNELXVAVLTGDGFAPTLDLS	241			
Dd	170	-----INLKDVFGAYSMDVITSTSFGVNIDSLNHPQD-----PVEVN	206			
QY	242	FKAIM-WDFV-PYERTMKRHRHEIPLTUGLA-VSRVGIELMEQKQOAVLGASDAQVDK-	297			
Dd	207	TKKULKFDFFLDPPF-----FSILLFPFLTPVPFBILNIWLFPFKVTOFFRKVSVERMKESR	260			
QY	298	-KDVOQR--DTLSLLVRANTAANLPESQKLSDBEVLQAISNLLPAGYETSGTVTWTMFHR	354			
Dd	261	LKDQKHREVFQLMINSONSKENDTHALSDELVAQSIIIFIAGVETTTSLSFLMYE	320			
QY	355	LSEDAVODKLBREICQI--DTDMPITDELNALPYLEAFVKESIRLPLDPPPSPYANRECLKD	412			
Dd	321	LATHPDVQQKLQEIBDATFNKALPTYDALVQMVELDWLNLETURLPYIAGRLERVCCKD	380			
QY	413	EDFTPLEAEPVIGRGGSVINERITKTGNWMPLPFINRSKFPIGEDABEEPRERKLEDVDT	472			
Dd	381	VE-----ISGVFIKPGVVVMVPTTLTRDOSLMPE-BEESPFRPFERSKNK	424			
QY	473	DSLNSIEAPYGHQASFISGPCACGMFAVAENKAEFLTVALRRVQPBPPI-SHPREYHILT	532			
Dd	425	DSIN----PTY-LPFGTGPNICGRPALNMKNMLALRVVLQNPSFCPC----KETQIPL	475			
QY	533	IISRPRIVGREKGGYQMRLOQVP	555			
Dd	476	KLNAOGHHIOPEK---PVLKVPE	495			

## RESULT 4

ADO47242

ID ADO47242 standard; protein; 501 AA.

XX ADO47242:

XX  
DT 29-III-20

DT 29-JUL-2004 (first entry)

DE Rabbit cytochrome P450 CYP3A6 SeqID 14.

XX dynamics docking; cytochrome P450; multispecific enzyme;  
KW protein coordinate data; common structural block; CSB; rabbit.

XX  
OS  
Oryctolagus cuniculus.

XX  
PN  
WO2004038655-A2.XX  
PD  
06-MAY-2004.XX  
PF  
28-OCT-2003: 2003WO-IB005134.XX  
PR 28-OCT-2002: 2002US-0421569P.

XX  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX  
PI Andre F, Delaforge M, Loiseau N;XX  
DR WPI; 2004-389974/36.

Sequence 503 AA;

us-10-066-007a-1.rag

Wed Dec 15 10:57:30 2004

PT defining 3D structure of CSBs, global constraints, selecting rotamers,  
 PT determining, optimizing 3D structures.  
 XX Claim 13; SEQ ID NO 14; 193pp; English.  
 XX This invention relates to a novel method for performing restrained  
 CC dynamics docking of one or more substrates on multi-specific enzymes.  
 CC Specifically, it refers to determining the three-dimensional structure of  
 CC active sites that are flexible and can adapt to different substrates i.e.  
 CC multispecific enzymes such as cytochrome P450. The present invention  
 CC describes identifying common structural blocks (CSBs) among members of  
 CC the family, aligning primary amino acid sequences, aligning the protein  
 CC as compared on the first alignment to obtain a second alignment, defining  
 CC 3D structure of protein CSBs, defining global constraints and selecting  
 CC rotamers, in order to determine the family of 3D model protein structures  
 CC and optimise models by dihedral angles. Accordingly, the method can be  
 CC used for screening, designing or identifying natural, unnatural or  
 CC substrate analogues, as well as inhibitors, activators or modulators of  
 CC the multispecific enzyme in question. In addition, it can determine the  
 CC effect of a first substrate on a second substrate, which can then be  
 CC applied to pharmaceutical products. Furthermore, the method can be used  
 CC to determine the oxidative modification of the substrate according to its  
 CC proximity to a haem molecule, for performing dynamic docking of the  
 CC metabolite either in the absence or presence of a second substrate in the  
 CC computed simulation and to compare the energy of the bound metabolite  
 CC relative to the energy of its parent metabolite bound, in order to  
 CC determine if the exit of the given metabolite from the enzyme is favoured  
 CC or not. This polypeptide sequence is a protein from the cytochrome P450  
 CC family of enzymes, given in an exemplification of the invention.  
 XX  
 SQ Sequence 501 AA;  
 Query Match 14.1%; Score 402; DB 8; Length 501;  
 Best Local Similarity 27.8%; Pred. No. 5.2e-26;  
 Matches 156; Conservative 88; Mismatches 200; Indels 118; Gaps 24;  
 QY 19 ASIAPFSLYLAPRSSLYNLOGPNHTY-FTGNFLDILSARTG--BEHAKYREKYGSTLR 75  
 DB 15 ASLVLLVLYGTHGLFKMGIPGTPPLPFGT---ILEYKRGWDFDIECRKYG---K 68  
 QY 76 FAGI--AGAPVNSTDPKVFHNM-KEAY-----DYPKPGMAARVLRATGDGVVTA 125  
 DB 69 MWGLFDGQPLWVITDPMIKTVLVKCYSVFTNRSFGPGVFMKKAISI-----S 120  
 QY 126 GEAKHRRHRIIMPISLAQAVKSMVPIFEKGMELVDKQMEDAEAKQAVGESAGEKATR 185  
 DB 121 DEDWKRVTLLSPTFTSGKLKEMLPITIAQYGDVILV-KNLQAEAKG----- 165  
 QY 186 LETGVVDKQWGRATLDVMAALGFDYKSDSLQNTNE-----LYVAFVGLTDGFAP 237  
 DB 166 ---KPVDDKEIFGAYSMVDVITGTSGVNDISLRNPQDFVKNVRLKFSF-----FDP 216  
 QY 238 TLDSPKALMDFVPVFTMKRHRHPIPTQGLAVSRVGIEMEQKQAVLGASDAQVDK 297  
 DB 217 LLLSI-----TLFPPL-----TPFEALHIS-MFPKDVMDFLKTSVEKIKDRLK 262  
 QY 298 KDVQRDILSLVRANIANIPESQKLSDEVLVAQISNLLFAGVETSGTSLTWMERHLS 357  
 DB 263 QK-RVDFQLMINSQNSKEIDSHKALDDIEVVAQSIILFAGVETSSLSLFINHLLAT 321  
 QY 358 DKAVQDKLREICQIDTDMPP-----TLDELNALPYLEAFVYESLRDPFPPYANRECKD 412  
 DB 322 HPDVQOKLQEE---IDTLLENKELATYDLVKMEYLDVWVNETRLRYPVLAGRLERVCKD 378  
 QY 413 EDFIPLABEVTRGDGSGVNEIVRTKGTWMLPLFNINRSKFIYEDAEERPERWLEDVT 472  
 DB 379 VD-----INGTFIPKGTITWMTYALHRDPOHWTE-DEFRPERFSKQNK 422  
 QY 473 DLSNS-IBAPYGHQASFTSGPRACPGWRFAVAMKAFILVTLRRVQFE-----PI--- 521  
 DB 423 DNINPIYHPFG-----AGPRNCLGMRFALMNIKALVRLMQNFSFKLCKETQVPLKLG 476  
 QY 522 ---ISHPEYHETTLIISPRIV 540

DB 477 KQGLQPEKPIVLKVSRDGII 498  
 RESULT 5  
 ADJ87521  
 ID ADJ87521 standard; protein; 485 AA.  
 XX AC ADJ87521;  
 XX 06-MAY-2004 (first entry)  
 XX Human cytochrome P450 3A4 protein.  
 XX cytochrome P450; crystallography.  
 OS Homo sapiens.  
 OS Unidentified.  
 XX WO2003102192-A1.  
 XX 11-DEC-2003.  
 XX 30-MAY-2002; 2002WO-GB002668.  
 XX 30-MAY-2002; 2002WO-GB002668.  
 XX (ASTE-) ASTEX TECHNOLOGY LTD.  
 XX Cosme J, Ward A, Vuillard L, Williams P, Hamilton B;  
 DR WPI: 2004-043119/04.  
 DR N-PSDB; ADJ87520.  
 XX Purifying a cytochrome P450, for NMR studies and high-throughput  
 PT screening methods to discover drugs, comprises suspending cells  
 PT expressing P450 molecule in a salt buffer, lysing the cells and providing  
 PT a high-salt-detergent lysate.  
 XX Claim 12; SEQ ID NO 8; 77pp; English.  
 XX The invention relates to a method of purifying a cytochrome P450 by  
 CC expressing in a host cell culture a cytochrome P450 molecule, recovering  
 CC the cells from the culture and suspending the cells in a salt buffer  
 CC having a conductivity of 12-110 mS/cm, lysing the cells and removing cell  
 CC debris to provide a high-salt lysate, adding to the lysate a detergent to  
 CC provide a high-salt-detergent lysate, and recovering the P450 from the  
 CC lysate. Methods of preparing and purifying cytochrome P450 proteins are  
 CC useful for X-ray crystallographic studies and crystallographic screening  
 CC of small molecules bound to P450, or for NMR studies and high-throughput  
 CC screening methods to discover drugs or analyze the interaction of drugs  
 CC with P450 molecules. This sequence corresponds to the human cytochrome  
 CC P450 3A4.  
 XX Sequence 485 AA;  
 SQ  
 Query Match 14.0%; Score 400; DB 8; Length 485;  
 Best Local Similarity 28.9%; Pred. No. 7.4e-26;  
 Matches 143; Conservative 83; Mismatches 176; Indels 92; Gaps 20;  
 QY 47 FTGNFLDILSARTG-----BEHAKYREKYGSTLRFAGIAGAPVNSTDPKVFHNM-KE 99  
 DB 24 FLGN---ILSYHKFGCMDFMECHKYKVGWGF---YDG--QOPVLAIITDPMIKTVLKE 75  
 QY 100 AYDY---PKPGMAARVLRATGDGVVTAEGEAKHRRHRIIMPISLAQAVKSMVPIFEK 156  
 DB 76 CYSVFTNRRPFGVFMKSA-----ISIADEDEWKLRLSLLSPTFTSGKLKEMVPIA 131  
 QY 157 MELVDKQMEDAEAKQAVGESAGEKATLEIF-EGVDVKQWGRATLDVMAALGFDYKSD 215  
 DB 132 DVLVRLNLRREA-----ETGKPVTLKDVFGAYSMVDVITSGVND 172  
 QY 216 SLQNKTNELYVAFVGLTDCGFAPTLDSFKAIM-WDFV-PYFRTMKRHE-IPUTQGLAV-- 270

```

Db 173 SLNPPQD-----PVENTKKLLRFDFLDPFFLSITVFPPELIPILVLCV 218
QY 271 -SRRVIGIELMOKKQAVLGSASDAQVKKQVQGRDILSLIVRANIAANLPSQKLSDEEV 329
Db 219 FPREVNTLRSVKR-----MKSERLEDTQHRVDFQLMIDSONSKETESHKALSDEL 273
QY 330 LAQISNLLFAGYETSSVLTWFMHRLSDKAVQDKLREICQI--DMDPTLDELNALPY 387
Db 274 VAQSIIFIFAGYETSSVLSFIMVELATHPDVQKQLEEDAVLPNKAPPTDYIVLQMEY 333
QY 398 LEAFVKESLRDPPSPYANRECLKDEPIPLAEPVIGRDSVINEVITKGTVMMLPLFN 447
Db 334 LDMVNETLRLFPPIAMRLERVCKKDE-----INGMFIPKGVVMIPSYA 378
QY 448 INRSKFIYGEDAEFRPRMLEDVTDLSLNS-IEAPYGHQASFISSGRACFGWFAVEMK 506
Db 379 LHRDP-KYWTPEPEKFLPERPSKKNKDNDPIVYTPFG-----SGPNCIGMFALMMK 431
QY 507 AFLFVTLRRVQPEP 520
Db 432 LALIRVLQNFSPK 445

```

## RESULT 6

ADO47238  
ID ADO47238 standard; protein; 503 AA.

AC ADO47238;

DT 29-JUL-2004 (first entry)

DE Human cytochrome P450 CYP3A3 SeqID 10.

KW dynamics docking; cytochrome P450; multispecific enzyme;

KW protein coordinate data; common structural block; CSB; human.

OS Homo sapiens.

PN WO2004038655-A2.

PD 06-MAY-2004.

PF 28-OCT-2003; 2003WO-IB005134.

PR 28-OCT-2002; 2002US-0421569P.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PA (CNRS ) CNRS CENT NAT RECH SCI.

PI Andre F, Delaforce M, Loiseau N;

PI WI; 2004-389974/36.

DR Designing 3 dimensional (3D) protein model, by identifying common structural blocks (CSBs) among family members, aligning sequence, defining 3D structure of CSBs, global constraints, selecting rotamers, determining, optimizing 3D structures.

PS Disclosure; SEQ ID NO 10; 193pp; English.

XX This invention relates to a novel method for performing restrained dynamics docking of one or more substrates on multi-specific enzymes. Specifically, it refers to determining the three-dimensional structure of active sites that are flexible and can adapt to different substrates i.e. multispecific enzymes such as cytochrome P450. The present invention describes identifying common structural blocks (CSBs) among members of the family, aligning primary amino acid sequences, aligning the protein as compared on the first alignment to obtain a second alignment, defining 3D structure of protein CSBs, defining global constraints and selecting rotamers, in order to determine the family of 3D model protein structures and optimise models by dihedral angles. Accordingly, the method can be used for screening, designing or identifying natural, unnatural or

CC substrate analogues, as well as inhibitors, activators or modulators of the multispecific enzyme in question. In addition, it can determine the effect of a first substrate on a second substrate, which can then be applied to pharmaceutical products. Furthermore, the method can be used to determine the oxidative modification of the substrate according to its proximity to a haem molecule, for performing dynamic docking of the metabolite either in the absence or presence of a second substrate in the computed stimulation and to compare the energy of the bound metabolite relative to the energy of its parent substrate bound, in order to determine if the exit of the given metabolite from the enzyme is favoured or not. This polypeptide sequence is a protein from the cytochrome P450 family of enzymes, given in an exemplification of the invention.

XX Sequence 503 AA;

Query Match 14.0%; Score 399; DB 8; Length 503;

Best Local Similarity 28.4%; Pred. No. 9.6e-26;

Matches 156; Conservative 83; Mismatches 182; Indels 128; Gaps 24;

```

QY 13 LAAFSW--ASIAFFSLYLAPRRS-SLYNLQG-PNHTNY-FTGNFLDILSARTG-----E 61
Db 6 LAMETWLLLAVALSVLLLYLGYTHSHGLFKKLGIPGPTLPFLGN---ILSYHKGFCMFDE 62
QY 62 EHAKYREKYGSTLRFAGTAGAPVLNSTDPKVFNVHM-KEAYD-----YKPGVAARVL 113
Db 63 CHKXGKVGWF---YDG--QQPVLAITDPMIKLVKCEYSVFTNRBPFGPVGPMKSAI 117
QY 114 RIATGDGVVTAEGEAHKHRRIMIPSLSAQAVKSVPIFLEKGMELVKMMEDAAEKDWA 173
Db 118 SI-----ADEEWKRLRSLLSPTFTSGKLKEWPPIAQVGDVLRNLRRE----- 162
QY 174 VGESAGEKKATRLTEGVVDKDWVGRATLDVNALAGPDYKSDSLQNKTNELYVAFVGLTD 233
Db 163 -----RETGKPVTLKDVFGAYSMDVITSSFGVNVDSLNNPD----- 200
QY 234 GFAPTLDSFKALM-WDFV-PYFRTWK-RRHEIPLTQGI-----AVSRVVG 275
Db 201 ---PLVENTKKLLRFDFLDPFFLSITVFPPELIPILVLCVPREVTNFKRAKRMKE 257
QY 276 IELMEQKQAVLGSASDAQVKKQVQGRDILSLIVRANIAANLPSQK-LSDEEVLQAIS 334
Db 258 SRLDETQKRV-----DFLQLMIDSHKNSKETESHKALSDELVAQSI 300
QY 335 NLLFAGYETSSVLTWFMHRLSEDKAVQDKLREICQI--DMDPTLDELNALPYLEAFV 392
Db 301 IFIFAGYETSSVLSFIMVELATHPDVQKQLEEDAVLPNKAPPTDYIVLQMEYLDWV 360
QY 393 KESLRDPPSPYANRECLKDEPIPLAEPVIGRDSVINEVITKGTVMMLPLFNIRSK 452
Db 361 NETLRLFPPIAMRLERVCKKDE-----INGMFIPKGVVMIPSYALHRDP 405
QY 453 FIYGEDAEFRPRMLEDVTDLSLNS-IEAPYGHQASFISSGRACFGWFAVEMKAFIFV 511
Db 406 -KYWTEPEKFLPERPSKKNKDNDPIVYTPFG-----SGPNCIGMFALMMKALIR 458
QY 512 TLRRVQPEP 520
Db 459 VLQNFSPK 467

```

## RESULT 7

ABU57260

ID ABU57260 standard; protein; 503 AA.

XX AC ABU57260;

XX DT 25-APR-2003 (first entry)

XX Human cytochrome P450 CYP3A4 protein.

DE Human; enzyme; transgenic; drug metabolism; behaviour; mouse;

XX Pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;

KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;





Matches	153;	Conservative	88;	Mismatches	195;	Indels	97;	Gaps	24;
QY	13	LAAPGW--ASIAFFSGLYAPRRS--SLYNLQG--PNHTNY--FTGNFLDILSARTG-----E	61						
Db	7	LAMETWLLLASLVLLLYLVTGSHGLFKKLGIPGTPLPFLGN---ILSYHKGFMPDME	63						
QY	62	EHAKYREKVGSLRPAFIAGAGAVILANSTDPKVNHYM--KEAYDY---PPGMAARVLRAT	117						
Db	64	CHKKGKVGWGF---VDG--QQPVLAITPDMTKTVLVREKYSVFTNRFPFGVGFMSA-	117						
QY	118	GDGVVTAEGEAHRRHRIMIPSLSAQAVKSMVPIFLEKGMELVDKMDAAEKDMAVGES	177						
Db	118	---ISIAEDBEWKRLRSLLSPFTTSGKLKEWPIITAOYGDVLVNLREAA-----	164						
QY	178	AGEKKAIRLET--EGVDVKDVGGRATLDNALAGFDYKSDSLQNKTNLYAVFVLTDGFA	236						
Db	165	---ETGKPVTLKDVFGAYGMDVITSTSGVNTDSLNNPD-----	201						
QY	237	PTLDSFKAIM--WDFV--PYFRMTK--RRHEIPTLTQGLAV--SRRVGIELEWQKKQAVLGS	290						
Db	202	PPVENTKKLLRDFDLDPFLSITVPFPLIPILEVLNICVFPREVNTFURKSVKR--M	256						
QY	291	SDQAVDKDVGQRDILSLVLRANITAAANIPESQKLSDEVLQAQISNLFAGYETSTVLTW	350						
Db	257	KESRLLEDYQKRVDPFLQMLIDSQNSKETESHKALSDLELVAQSIIFIFAGYETSSVLSF	316						
QY	351	MFHRLSDEKAVQDKLREIICOI--DTDMPTLDELNALPYLEAFVKESLRILDPSPFYANRE	408						
Db	317	IMVELATHPDVQOQLQEEIDAVLNKAPPTDYTLQMEYELDMVNYETRLRFPFIAMRLRV	376						
QY	409	CLKDEDFIPLAEPVIGRDSGVINEVRITKGTMMPLPLFININRSKFICYGDAEEFPRPWL	468						
Db	377	CKKDVE-----INGWFPGVGVVMPSPVALHRDP-KWYTEPEKELPERFS	420						
QY	469	EDVTDLSLNS--IEAPYGHCASTISGRACFGWRFAVAKPFLFTLRRVQEP	520						
Db	421	KKNKDNDIPYITPFG-----SGFRNCIGMRFMNMKILARLVQLNQSFKP	467						
RESULT	9								
ADLI18578	ID	ADLI18578 standard; protein; 503 AA.							
XX	AC	ADLI18578;							
XX	DT	06-MAY-2004 (first entry)							
XX	XX	Human wild-type cytochrome P450 CYP3A4 protein.							
XX	DE	drug metabolic activity; CYP3A4; human; wild-type; cytochrome P450; SNP;							
XX	KW	single nucleotide polymorphism.							
XX	OS	Homo sapiens.							
XX	XX								
XX	Key	Location/Qualifiers							
FT	FT	Misc-difference 185							
FT	FT	/note= "Thr may be substituted for Ser due to single							
FT	FT	nucleotide polymorphism (SNP) in coding DNA"							
FT	FT	Misc-difference 293							
FT	FT	/note= "Leu may be substituted for Pro due to single							
FT	FT	nucleotide polymorphism (SNP) in coding DNA"							
FT	FT	Misc-difference 363							
FT	FT	/note= "Thr may be substituted for Met due to single							
FT	FT	nucleotide polymorphism (SNP) in coding DNA"							
XX	XX								
XX	XX	JP2004000004-A.							
XX	XX								
XX	PD	08-JAN-2004.							
XX	XX								
XX	XX	23-APR-2002; 2002JP-00120246.							
XX	XX								
XX	XX	23-APR-2002; 2002JP-00120246.							
XX	XX								

(KOKU-) KOKURITSU IYAKUJIN SHOKUHIN EISEI KENKYU.  
(IYAK-) IYAKUJIN FUKUSAYO HIGAI KYUSAI KENKYU SH.  
WPI; 2004-085203/09.  
N-PSDB; ADL18577.

Testing the drug metabolic activity of CYP3A4 such as testosterone hydroxylase activity by detecting the CYP3A4 gene polymorphisms in which the drug metabolic activity of CYP3A4 is reduced.

Disclosure; SEQ ID NO 2; 48pp; Japanese.

The invention relates to a novel method for testing the drug metabolic activity of CYP3A4 by detecting the CYP3A4 gene polymorphisms in which the drug metabolic activity of CYP3A4 is reduced. The method of the invention may be useful for testing the possibility of a side effect resulting from a drug metabolised by CYP3A4 in a subject. The current sequence is that of the human wild-type cytochrome P450 CYP3A4 protein of the invention.

SEQ ID NO 2; 48pp; Japanese.

Query Match 14.0%; Score 398.5; DB 8; Length 503;  
Best Local Similarity 28.7%; Pred. No. 1.1e-25;  
Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24

13 LAAPSW--ASTAFSLYLAPRS--SLYNLQ--PNHTNY--FTGNFUDILSARTG-----E 61  
7 LAMETWLLAVSLVLLVLYGTHSHGLFKLGPQTPLPFLGN---ILSYHKGFQWFDME 63

62 EHAKYREKYGSTLAFAGIAGAPVLNSTDPKVENHVM--KEAYDY---PKPGMAARVLRAT 117  
64 CHKKYGVWGF--YDG--QQPVLAITDPMIKTVLKECVSVFTNRRPFGVPGMKGA- 117

118 GDGVVTAEGEAKHRRIMIPSLSAQAVKSMVPIFEKGMELVDKMMEDAAEKDVAVGES 177  
118 ---ISIAEDEEWKELRSILSPFTTSGKLKEMVPIIAQYGDVLVRLRREA----- 164

178 AGEKKAATLET--EGVDYKDWVGRATLDVMAIAGDYKSDSLONKTNELYVAFVLTDGFA 236  
165 -----ETGKPVTLKDFVFGVSMVITSTSGVNIDSUNNPQD----- 201

237 PTLDSFKALM--WDFV--PYFRMK--RRHEIPLTQGLAV---SRRVGIELMEQKQAVLGSA 290  
202 PFVENTKKLRFDFLDFPFLSITVFPFLIPIELVNLICVPREVTNLRKSVKR-----M 256

291 SDQAVDKDVQGRDILSLIVRAIANLPSQKLSDEEVLAAQISNLLFAGYETSTVLTV 350  
257 KESRLDQTKHRVDFLQLMIDSNQSKTESHKALSLELVAQSIIFIFAGYETTSSVLSF 316

351 MFHRLSDKAVQKLRKEIQI--DTPMTPLDELNALPYLEAFVKESLSRLDPPSPYANRE 408  
317 IMVELATHPDVQQLQEEIDAVLPNKAPFYDVLQMEYLDVMVNEVRLRFLPIANRLRV 376

409 CLKDEDPIPAEPVIGRDGSGVINVEIRITKGTVMVLPNFNIRSKFIYGEDAEERPRWL 468  
377 CKKDVE-----INGMFIPKGVWVWIPSYALHRDP--KYWTEPEKFLPERFS 420

469 EDTVDSLNS--IEAPYGHQASFISGRACFGWRFAVAEMKAFIPLVTLRRVQPEP 520  
421 KKKNDMIDPVIYTPFG-----SGPRNCIGWRPALMMKALIRVLQNFQSFKP 467

RESULT 10.  
ADL18672  
ID ADL18672 standard; protein; 503 AA.  
XX  
AC ADL18672;  
XX  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Human cytochrome P450 enzyme 3A4 protein.  
XX

RESULT 10
ADL18672
ID ADL1
XX
AC ADL1
XX
DT 17-J
XX
DE Huma
XX







PT Antibody recognising human derived cytochrome P450A4 - allows specific  
 PT detection of cytochrome P450 species in humans.  
 PS Example 1; Page 10-12; 13pp; Japanese.

XX The present sequence is the human derived cytochrome (HDC) P450A4, which  
 CC was obtained from a commercial cDNA library. Yeast were transfected with an  
 CC expression vector contg. the HDC cDNA, cultured and then disrupted to  
 CC give a microsomal fraction. The HDC was purified from the fraction, and  
 CC used to immunise and sensitise a mammal. Blood was drawn from the mammal,  
 CC and an anti-HDC antibody isolated. The antibody obtd. recognises HDC,  
 CC P450A4, partic. at a serum dilution rate of 1:10000, and is  
 CC substantially without cross reaction to other HDC P450 spp

XX Sequence 503 AA;

Query Match 13.9%; Score 396.5; DB 2; Length 503;

Best Local Similarity 28.7%; Pred. No. 1.6e-25;

Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;

QY 13 LAAFSW--ASIAFFSLYLAPRRS--SLYNLQ--PNHTNY--FTGNFLDILSARTG-----E 61  
 DB 7 LAMETWLLAVSLVLLYLYGTHSHGLPKKLGIPGTPLPFLGN---ILSYHKGFCMPDME 63  
 QY 62 EHAKYREKYGSTLRPAGTAGAPVLNSTDPKVFNVHM-KEAYDY---PKPGMAARVLRAT 117  
 DB 64 CHKKGKVGWGF---YDG--QQFVLAITDPMIKTVLKECVSVFTNRRPFGVGMKSA- 117  
 QY 118 GDGVVTAEGEAHKKRRIRIMISLSAQAVKSMVPFIFLEKGMELVDKMDAEAKDMVGES 177  
 DB 118 ---ISIADEDEWKLRLSLLSFTFTSGKLEKMPPIAQGVDLVNLNREA----- 164  
 QY 178 AGEKATRLT-EGVDVKDWVGRATLDVMAAGFDYKSDSLQNKTNELYVAFVGLTDGFA 236  
 DB 165 -----ETGKPVTLKDFEVSMDVITSTSGVNIIDSLNPNQD----- 201  
 QY 237 PTLDSFKAIM-WDFV-PYFRTMK-RRHEIPLTQGLAV---SRRVGIEMEQKQAVLGSA 290  
 DB 202 PFVENTKKLRFDFDFFLSITVFPFLIPLEVLNLCVFPREVTNFRKSVKR-----M 256  
 QY 291 SDQAVDKDQVGRDILSLLVRANIAANLPESQKLSDEVLQAQISNLLFAGYETSSVLTW 350  
 DB 257 KESRLEDTQKRVDFLQLMIDSONSKETESHKALSDELVQAQSIIFIFAGYETSSVLSF 316  
 QY 351 MFHRLSEDKAVODKLRREICQI--DTDMPILDELNALPYLEAFVKESLRDPPSPYANRE 408  
 DB 317 IMYELATHPDVQOKLQBEIDAVLPNKAPPTDYTLQMEYLDVMVNETLRLFPFIAMRLRV 376  
 QY 409 CLKDEDFIPLAEPVIGRDGVSINEVRITKGTVMWMLPLFNINRSKFIYGEDAEFRERWL 468  
 DB 377 CKKDVE-----INGMFIPKGVWVNIYSYALHRDP-KYWTBPEKFLPERFS 420  
 QY 469 EDVTDLSNS-IEAPYGHQASFIQPRACFGWRFAVAMKAFLEVTLLRRVQFEP 520  
 DB 421 KKNKNDIDPIYTPFG-----SGPRNCIGMRPALMMKALIRVLQNSFKP 467

RESULT 15

AAR93170

ID AAR93170 standard; protein; 503 AA.

XX AAR93170;

AC AAR93170;

DT 11-OCT-1996 (first entry)

XX Human cytochrome P450 molecular species 3A4 protein.

KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;  
 KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;  
 KW evaluation; safety; fusion protein; metabolite; detoxification;  
 KW carcinogenic.

XX Homo sapiens.

OS

XX JF08056695-A.

XX 05-MAR-1996.

XX 15-JUL-1994; 94JP-00164184.

XX 20-JUL-1993; 93JP-00201120.

XX 30-JUL-1993; 93JP-00208279.

XX 17-JUN-1994; 94JP-00136053.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX WPI; 1996-182311/19.

XX N-PSDB; AAT28383.

XX Novel method for the evaluation of the safety of a cpd. - using a human  
 PT cytochrome P450 and yeast NADPH reductase to determine whether the  
 PT analyte cpd. is detoxified or metabolised to a carcinogen.

XX Example 1; Page 24-26; 74pp; Japanese.

XX This is the amino acid sequence of the human cytochrome P450 molecular  
 CC species 3A4 protein. The corresp. gene was amplified from a human liver  
 CC derived cDNA library as 2 fragments of 0.6 and 0.9 kb using primers  
 CC AAT26933-6. The prod. was cloned into the yeast expression vector PAAHEN  
 CC to generate plasmid p3A4 for prodn. of the cytochrome only or into the  
 CC vector PAHRR to generate the plasmid p3AAR for co-prodn. with the yeast  
 CC NADPH-P450 reductase. The sequence is placed under control of the yeast  
 CC ADH gene promoter and terminator. The vectors are used in a method for  
 CC evaluating the safety of a cpd. by reacting the test cpd. with  
 CC recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),  
 CC 2C9 (AAT28381), 2B1 (AAT28382), 3A4 or their variants (AAT28384-98)  
 CC together with yeast NADPH-P450 reductase (either as a fused protein or as  
 CC a cell extract) and analysing the resultant metabolite. The cpd. is  
 CC considered "safe" if it is detoxified or not rendered carcinogenic or  
 CC "unsafe" if it is not detoxified or is metabolised to a carcinogenic cpd

XX Sequence 503 AA;

Query Match 13.9%; Score 396.5; DB 2; Length 503;

Best Local Similarity 28.7%; Pred. No. 1.6e-25;

Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;

QY 13 LAAFSW--ASIAFFSLYLAPRRS--SLYNLQ--PNHTNY--FTGNFLDILSARTG-----E 61  
 DB 7 LAMETWLLAVSLVLLYLYGTHSHGLPKKLGIPGTPLPFLGN---ILSYHKGFCMPDME 63  
 QY 62 EHAKYREKYGSTLRPAGTAGAPVLNSTDPKVFNVHM-KEAYDY---PKPGMAARVLRAT 117  
 DB 64 CHKKGKVGWGF---YDG--QQFVLAITDPMIKTVLKECVSVFTNRRPFGVGMKSA- 117  
 QY 118 GDGVVTAEGEAHKKRRIRIMISLSAQAVKSMVPFIFLEKGMELVDKMDAEAKDMVGES 177  
 DB 118 ---ISIADEDEWKLRLSLLSFTFTSGKLEKMPPIAQGVDLVNLNREA----- 164  
 QY 178 AGEKATRLT-EGVDVKDWVGRATLDVMAAGFDYKSDSLQNKTNELYVAFVGLTDGFA 236  
 DB 165 -----ETGKPVTLKDFEVSMDVITSTSGVNIIDSLNPNQD----- 201  
 QY 237 PTLDSFKAIM-WDFV-PYFRTMK-RRHEIPLTQGLAV---SRRVGIEMEQKQAVLGSA 290  
 DB 202 PFVENTKKLRFDFDFFLSITVFPFLIPLEVLNLCVFPREVTNFRKSVKR-----M 256  
 QY 291 SDQAVDKDQVGRDILSLLVRANIAANLPESQKLSDEVLQAQISNLLFAGYETSSVLTW 350  
 DB 257 KESRLEDTQKRVDFLQLMIDSONSKETESHKALSDELVQAQSIIFIFAGYETSSVLSF 316  
 QY 351 MFHRLSEDKAVODKLRREICQI--DTDMPILDELNALPYLEAFVKESLRDPPSPYANRE 408  
 DB 317 IMYELATHPDVQOKLQBEIDAVLPNKAPPTDYTLQMEYLDVMVNETLRLFPFIAMRLRV 376  
 QY 409 CLKDEDFIPLAEPVIGRDGVSINEVRITKGTVMWMLPLFNINRSKFIYGEDAEFRERWL 468



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2004, 14:35:44 ; Search time 45 Seconds  
(without alignments)  
1190.950 Million cell updates/sec

Title: US-10-066-007A-1  
Perfect score: 2852  
Sequence: 1 MFILVLLTGALGAFAFWAS.....RIVGREKEGYQMRLOVKPVE 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	419	14.7	593	2 F86441	probable cytochrom
2	406	14.2	503	2 JC4702	cytochrome P450 3A
3	404	14.2	503	2 S14275	steroid 6beta-mono
4	402	14.1	501	2 A34236	cytochrome P450 3A
5	400	14.0	526	2 T02191	cytochrome P450 ho
6	399	14.0	504	2 A29410	cytochrome P450 g
7	398.5	14.0	503	2 A29815	cytochrome P450 3A
8	398	14.0	503	2 S50211	cytochrome P450 3A
9	395.5	13.9	503	2 S28168	nifedipine oxidase
10	394	13.8	504	2 A60564	cytochrome P450 3A
11	385.5	13.5	503	2 JX0062	cytochrome P450 3A
12	385	13.5	502	1 A34101	cytochrome P450 3A
13	379	13.3	501	2 A29487	cytochrome P450 3A
14	370	13.0	504	2 A22631	cytochrome P450 3A
15	368	12.9	504	2 S50892	cytochrome P450 3A
16	365	12.8	520	2 H84663	probable cytochrom
17	363.5	12.7	511	2 T00864	probable cytochrom
18	362.5	12.7	523	2 B96662	probable cytochrom
19	359.5	12.6	511	2 S64472	cytochrome P450 4B
20	357	12.5	502	2 JX0334	cytochrome P450 3A
21	355.5	12.5	512	2 A96695	hypothetical prote
22	350.5	12.3	1054	1 A69975	NADPH:ferrihemopro
23	350	12.3	526	2 JC4533	cytochrome P450 4F
24	349.5	12.3	511	1 O4HUB1	cytochrome P450 4B
25	349.5	12.3	524	2 S27223	cytochrome P450 4F
26	346.5	12.1	511	1 B40164	cytochrome P450 4B
27	346.5	12.1	524	2 JC7594	cytochrome P450 en
28	346.5	12.1	524	2 JC7598	cytochrome P450 en
29	346	12.1	537	2 T02450	Probable cytochrom

30 345.5 12.1 520 2 T24778 hypothetical prote  
31 345 12.1 516 2 T00514 cytochrome P450 ho  
32 344 12.1 520 1 A46661 leukotriene B4 ome  
33 343 12.0 504 2 A25222 cytochrome P450 3A  
34 340.5 11.9 524 2 T09944 probable cytochrom  
35 340 11.9 517 2 T02192 probable cytochrom  
36 339.5 11.9 508 2 T16980 probable cytochrom  
37 338.5 11.9 503 2 JC7627 cytochrome P450 3A  
38 338.5 11.9 515 2 JC8026 cytochrome P450 en  
39 337.5 11.8 524 2 T09999 cytochrome P450 en  
40 337 11.8 520 2 S45702 cytochrome P450 -  
41 336.5 11.8 490 2 T00404 leukotriene-B4 20-  
42 336 11.8 522 2 JC4532 probable cytochrom  
43 334.5 11.7 509 2 T24785 cytochrome P450 4F  
44 334.5 11.7 576 1 H71414 hypothetical prote  
45 333.5 11.7 497 2 S52097 probable cytochrom  
cytochrome P450III

ALIGNMENTS

RESULT 1

F86441  
probable cytochrome P450 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: F86441  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, A.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maici, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: F86441  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-593 <STO>  
A:Cross-references: UNIPROT:Q9C6S0; GB:AB005172; NID:G1136728; PIDN:AAG31309.1; GSPDB: C:Genetics:  
A:Map position: 1  
C:Superfamily: pea cytochrome P450 CYP97; cytochrome P450 homology  
C:Keywords: heme; iron; metalloprotein  
F:514/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 14.7%; Score 419; DB 2; Length 593;  
Best Local Similarity 26.4%; Pred. No. 8.5e-20;  
Matches 134; Conservative 95; Mismatches 179; Indels 100; Gaps 19;

QY 50 NFDILSARTG--EEHAKYREKYGST-----LRFAGI-----AGAPVLN 86  
Db 93 NVLDFMDTGDSDQDYPKVEAKGSIQAVNEAFFIPLYELFTYGGIFLTFGPKSFLI 152  
QY 87 STDPKVFNVHMK--AYDYPKPGMAARVLRATGDGVVTAEGAAKRRHRRIMPSLSAQAV 145  
Db 153 VSDPSIAKHILKDNAXAYSK-GILAEILDFVMGKLIPADGEIWRERRRAIVFALHQYV 211  
QY 146 KSMVPIELEKGMELVDQMDEDAEKDMVAGESAGEKATRETEGVVDKDWGGRATLDVM 205.  
Db 212 AAMLSLFGESDRUCQKL--DAA-----ALKGEEVEMESLFRSLTLDII 253  
QY 206 ALAGFDYKSDSLQKNTNELYVAVFGLTGDGPATLDSFKALIMDFVPVFRMKRRHRIPLT 265  
Db 254 GKAVFNVDPSLNDTGVIEAVTVLREADRVSPIP--VND-IFWKDIS-----PRQ 305  
QY 266 QGLAVSRVCIEMEQKQAVLSAQDQVKKDQV-----GRD--ILSLIVRANIAA 316  
Db 306 RKVATSLKLNITLDD-----LIATCKRMVEEBELQFHEEYMNWRDPSILHFL----- 353







cytochrome P450 3A6 (version 2) - rabbit  
N/Alternate names: cytochrome P450 3c  
N/Contains: oxidoreductase (EC 1.-.-.-)  
C/Species: Oryctolagus cuniculus (domestic rabbit)  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C/Accession: A34236  
R/Potenza, C.U.; Pendurthi, U.R.; Strom, D.K.; Tukey, R.H.; Griffin, K.J.; Schwab, G.E.;  
J. Biol. Chem. 264, 16222-16228, 1989  
A/Title: Regulation of the rabbit cytochrome P-450 3c gene. Age-dependent expression and  
A/Reference number: A34236; MUID:89380226; PMID:277787  
A/Accession: A34236  
A/Molecule type: mRNA  
A/Residues: 1-501 <POT>  
A/Cross-references: UNIPROT:P11707; GB:J05034; NID:G164829; PID:AAA31178.1; PID:G164830  
C/Genetics:  
A/Gene: CYP3A6  
C/Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homolog  
C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F/300-462/Domain: cytochrome P450 homolog <P4>  
F/440/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 14.1%; Score 402; DB 2; Length 501;  
Best Local Similarity 27.8%; Pred. No. 8.8e-19;  
Matches 156; Conservative 88; Mismatches 200; Indels 118; Gaps 24;

QY 19 ASIAFTSLYLAPRRSSLYNLQGNHHNY-FTGNFLDILSARTG--ZEHAKEYBEKYGSTLR 75  
DB 15 ASLVLLYLYGTSTHGKFKMGIGPGLPFIQT---ILEYRGIMDFDIECRKKY--K 68  
QY 76 FAGI--AGAPVLNSTDPKFNHVM-KEAY-----DYPKPGMAARVLRIATGCGVTAEE 125  
DB 69 MWGLFDGROPMLVITDPDMIKTVLKECVSVFTNRRSGFPGVFMKKAVERI-----SE 120  
QY 126 GEAFKEHRRIMIPLSAQAVKSMVPIFLKGMELVDKMMEDAAEKDMAGESAGEKATR 185  
DB 121 DEWKVRVRLTSLFTFTSGKLKEMLPPIAQYGVDLV-KNLRQEAERG----- 165  
QY 186 LETEGVDVQDWGRATLDVNALAGFDYKDSLQNKTE-----LYVAFVGLTDGEPAP 237  
DB 166 ---KPVDLKEIFGAYSMDVITGTFSGWNIDSLRNPQDPFVKNVRLLKESF-----FDP 216  
QY 238 TLDSFKAIMWDFVPYFRTMKRHEIPLTGLAVSRVRVGIELMEQKQAVLGSASQAVDK 297  
DB 217 LLSLI-----TUFFPL-----TFIFALHIS-MFPKDVMDFLKTSVEKIKDDRLLDK 262  
QY 298 KDVQGRDIILSLAVRANIAANLPESOKLSDVEEVAQISNLLFAGYETTSVLTWMPHLSE 357  
DB 263 QK-RRVDFQLQMINSQNSKEIDSHKALDDIEVVAQSIIILFAGYETTSVLSFMILLAT 321  
QY 358 DKAVQDKLBEETCQIDTMP-----TLDELNALPYLEAFVZESRLDPPSPVANRECLKD 412  
DB 322 HPDVOOKQEE---IDTLFNKELATYDTLVKMEYLDVMVNETLRLPYIAGRLERVCKKD 378  
QY 413 EDFIPLAEPIVGRDGSVINEVRI TGTVMVLMPLFNINESKFIYGEDAEFFPERWLEDVT 472  
DB 379 VD-----INGTIFPKGTIVMMFTYALHRDPQHWTB-PDEFPRPFSKKNK 422  
QY 473 DLSNS-IEAPYGHQASFISSGPACFGWRFAVEMKAFVLTLRRVQFE-----PI--- 521  
DB 423 DNINPYIYHPFG-----AGPRNCLGMRFALMNKIALVRLMQNFSFKLCKETQVFLKLG 476  
QY 522 ---ISHPEYEHITLIIISPRIV 540  
DB 477 KQGLLQPEKPIVLKVVSRDGI 498

RESULT 5  
T02191  
cytochrome P450 homolog F14M4.21 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 05-Mar-1995 #sequence\_revision 05-Mar-1999 #text\_change 09-Jul-2004  
C/Accession: T02191; D84909  
R/Kounisley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ro

```

submitted to the EMBL Data Library, September 1998
A>Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
A>Reference number: Z14609
A>Accession: T02191
A>Status: translated from GB/EMBL/DBJ
A>Molecule type: DNA
A>Residues: 1-526 <ROU>
A>Cross-references: UNIPROT:Q9ASR3; EMBL:AC004411; NID:g3522932; PID:g3522945
A>Experimental source: cultivar Columbia
R>Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujiki,
M.; Koo, H.; Moffat, K.C.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.;
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A>Reference number: A84420; MUID:20083487; PMID:10617197
A>Accession: D84909
A>Status: preliminary
A>Molecule type: DNA
A>Residues: 1-526 <STO>
A>Cross-references: GB:AE002093; NID:g3522945; PIDN:AAC34227.1; GSPDB:GN00139
C>Genetics:
A>Gene: At2g46960; F14M4.21
A>Map position: 2
A>Introns: 93/1; 169/3; 254/2; 375/3
C>Superfamily: human cytochrome P450 CYP4B1; cytochrome P450 homology
C>Keywords: heme; iron; metalloprotein
F:322-493/Domain: cytochrome P450 homology <P45>
F:471/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match          14.0%; Score 400; DB 2; Length 526;
Best Local Similarity 23.8%; Pred. No. 1.3e-18;
Matches 140; Conservative 109; Mismatches 231; Indels 114; Gaps 21;

Qy 1 MFILVLITGALGAFSAFWSIASIAFFPSIYLAPRSSLYNLOCPNHNTYFTGNF----- 51
Db 9 LVLILILI---IGLRIPKAFMILVWHFVLTRLLKNGQIGSPNY-RIFYGNLSLSEIKMKKE 64

Qy 52 --LDLSARTGE-----EHAKREKYKYSTLRFAGIAGAPVLNSTDPKVFNHVKBYD 102
Db 65 SHLISLDPSNDITPRLPHYQKMSQGETFLYWN-GTEPRICISDPELAKTMLSNLKG 123

Qy 103 -YKPGQMAARVLRIATGDGVGTAEGEAHRRHRIIPSLSAQAVSMVPIFEKGMELVD 161
Db 124 FFVKSKARPEAVKLVGSKGLVFTGADWVRHRRILNPAFISIDRLKIMTVMVVDCTL--- 179

Qy 162 KMWEDAAEKDVMGESAGSEKKATRLTEGEVDVQDWVG-----RATLDVMALAGFDYKSDSL 217
Db 180 KMLSE-----WKESKTEKETHPKIKCMNEEFQRLTADIATSAGF----- 221

Qy 218 QNKTNELYVAVGLTDGFAPTLDSEFKAIMWDFVPYFRYMKRHEIPLTQGLAVSRRVGIE 277
Db 222 -----SSYVEGIEVFRSQM-ELKRCYTTSLNQVSIPTQYLTPTSNIRW 265

Qy 278 LMEOCKQAVLGASDAQDKVDQGRDILSLVRANIANYPESOKLSDSEVLQISNLL 337
Db 266 KLERKMDNISIKRIISRLQSKSDYGDGLGILLKAYNTEG--KERKMGIEETIHECRATFF 323

Qy 338 FAGYETSTVLTMWFHRLSEDAVKDKLREET---C--QIDTDMPTLDLNLALP---YL 388
Db 324 FGGHETTSNLLATWTTMLLSLHODWQEKLEEEIFKECGEKTPDSETFSKLKLKPYVFLQM 383

Qy 389 EAFVKESLRLDPPSPYANRECLKDSDFTPLAEPIVGRGCV---INEVRIKGTWMLPL 445
Db 384 NMVIRSESLYGP-----VSALAREASVNIKGLDLEIPKGTITWVIFL 425

Qy 446 FMINRSKFTYGSDAEEFPERMLEVDVTDLSNLSIEAPYGHQASFISGPRACGWRFAVEM 505
Db 426 LKXWSDKTLWGSADAKFNPMPFANGSVRAAHFNA-----LLAFSGVPRACIGQNFVIEA 481

Qy 506 KALFVTVLRRVQPEFIISHPEYEH--ITLISRPRIVRGREGYQMRLOVKPVE 557
Db 482 KTVLTMILQRFIFSLCD--EYKHTFVDNVTIQPO-----XCLPWLQPLE 525

```

RESULT 6

A29410  
cytochrome P450, glucocorticoid-inducible, hepatic - human  
N/Alternate names: cytochrome P450 3A3; cytochrome P450 HLP  
N/Contains: oxidoreductase (EC 1.-.-.-)  
C/Species: Homo sapiens (man)  
C/Date: 31-Mar-1989 #sequence revision 31-Mar-1989 #text\_change 09-Jul-2004  
C/Accession: A29410; A25170; A25450  
R/Molowa, D.T.; Schuetz, E.G.; Wrighton, S.A.; Watkins, P.B.; Kremers, P.; Mendez-Picon,  
Proc. Natl. Acad. Sci. U.S.A. 83, 5311-5315, 1986  
A/Title: Complete cDNA sequence of a cytochrome P-450 inducible by glucocorticoids in hu  
A/Reference number: A94111; MUID:86259780; PMID:3460094  
A/Accession: A29410  
A/Molecule type: mRNA  
A/Residues: 1-504 <MOL>  
A/Cross-references: UNIPROT:P05184; GB:M13785; NID:G181353; PIDN:AAA35742.1; PID:G181354  
R/Watkins, P.B.; Wrighton, S.A.; Maurer, P.; Schuetz, E.G.; Mendez-Picon, G.; Parker, G.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6310-6314, 1985  
A/Title: Identification of an inducible form of cytochrome P-450 in human liver.  
A/Reference number: A25170; MUID:85298342; PMID:3898085  
A/Accession: A25170  
A/Molecule type: protein  
A/Residues: 2-21 <WAT>  
C/Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C/Keywords: chromoprotein; electron transfer; heme; iron; liver; metalloprotein; monooxy  
F:303-465/Domain: cytochrome P450 homology <P45>  
F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 14.0%; Score 399; DB 2; Length 504;  
Best Local Similarity 28.4%; Pred. No. 1-4e-18;  
Matches 156; Conservative 83; Mismatches 182; Indels 128; Gaps 24;

QY 13 LAAFSW--ASTAFPSLYLAPRRS--SLYNLOG-PNHNTY-FTGNFLDILSARTG-----E 61  
DB 7 LAMETWLLAVSLVLLVLYGTHSHGLFKLGIPTPLPLGN---ILSYHKGFCMFDM 63

QY 62 BHAKYREKYGSTLRFAGIAGAPVLSNTPDKVFNHVM-KEAYD-----YPKPGMAARVL 113  
DB 64 CHKKYKVGWGF---YDG--QQPVLAITDPDMIKLVKVCYSVFTNREPFGVFKMSAI 118

QY 114 RIATGDGVVTAEGEAHKEHRRIMIPSLSAQAVKSMVPIFILEKGMELVDKMDAAEKDMA 173  
DB 119 SI-----ADESEWKEKRLSLLSPTFTSGKLEKMPVIAQYGVLVNLRRE----- 163

QY 174 VGESAGEKATRLTEGVVDVQVQVGRATLDYMALAGDYKSDSLONKTNELYVAFVGLTD 233  
DB 164 -----RETGKPVTLKDVFGAYSMOVTSSFGVNVDSLNNPQD----- 201

QY 234 GFAPLTDGFKALM-WDFV-PYFRFMK-RRHEIPTQGL-----AVSRVVG 275  
DB 202 ---PLVNTKLLRFDLDFLSTVTPFPFLIPVLEVLNLCVFPREVTFNLRKAVKMK 258

QY 276 TELMEQKQAVLGSADQAVDKQVQGRDILSLVRANIANLPSQK-LSDEVLQAQIS 334  
DB 259 SRLSDTKHRV-----DFLQLMIDSHKNSKETESHKALSDLEVAQSI 301

QY 335 NLLPAGYETSSVLTWMPHRLSEDAVDKLRREICQI--DTPMTLDELNALPYLEAFV 392  
DB 302 IFIFAGYETSSVLSFIMWEATHDPVQVKQLEIDAVLPNKAPPTVTVLQMEYLDWV 361

QY 393 KESRLDPPSPYANRECLDKEDFDIPLAEPVIGRDSGVINEVRIKGTWMLPLFNINRSK 452  
DB 362 NETLRLFPIMRLERVCVKDVE-----INGMFIPKGVWVWVIPSVALHRDP 406

QY 453 FIYGDAEPRFRWLEDVTDLSNS-IEAPYGHQASFTSGPRACFGMRFAVEMKALFV 511  
DB 407 -KYWTEPEKFLPERESKKONIDPIYITPFG-----SGPRNCIGVRFALNNKMLALIR 459

QY 512 TLRRVQFEP 520  
DB 460 VLQNFSEFKP 468

RESULT 7

A29815  
cytochrome P450 3A4 nifedipine oxidase (EC 1.14.14.-) - human  
N/Alternate names: cytochrome P450 (PCN1); cytochrome P450-HM1  
C/Species: Homo sapiens (man)  
C/Date: 30-Sep-1989 #sequence revision 30-Sep-1989 #text\_change 09-Jul-2004  
C/Accession: A29815; S16900; A25177; A32199; PX0012; S03851  
R/Gonzalez, F.J.; Schmid, B.J.; Umeno, M.; McBride, O.W.; Hardwick, J.P.; Meyer, U.A.; G  
DNA 7, 79-86, 1988  
A/Title: Human P450PCN1: sequence, chromosome localization, and direct evidence through  
A/Reference number: A29815; MUID:88195781; PMID:3267210  
A/Accession: A29815  
A/Molecule type: mRNA  
A/Residues: 1-503 <GON>  
A/Cross-references: UNIPROT:P08684; GB:M18907; NID:G181373; PIDN:AAA35745.1; PID:G181374  
R/Spurr, N.K.; Gough, A.C.; Stevenson, K.; Wolf, C.R.  
Hum. Genet. 81, 171-174, 1989  
A/Title: The human cytochrome P450 CYP3 locus: assignment to chromosome 7q22-qter.  
A/Reference number: S16900; MUID:89108438; PMID:2563251  
A/Accession: S16900  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-391, 'W', 393-503 <SP2>  
A/Cross-references: EMBL:X12387; NID:G35910; PIDN:CAA30944.1; PID:G35911  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1988  
R/Beaune, P.H.; Umbenhauer, D.R.; Bork, R.W.; Lloyd, R.S.; Guengerich, F.P.  
Proc. Natl. Acad. Sci. U.S.A. 83, 8064-8068, 1986  
A/Title: Isolation and sequence determination of a cDNA clone related to human cytochrome  
A/Reference number: A25517; MUID:87041402; PMID:3464943  
A/Accession: A25517  
A/Molecule type: mRNA  
A/Residues: 1-391, 'W', 393-503 <BEA>  
R/Bork, R.W.; Muto, T.; Beaune, P.H.; Srivastava, P.K.; Lloyd, R.S.; Guengerich, F.P.  
J. Biol. Chem. 264, 910-919, 1989  
A/Title: Characterization of mRNA species related to human liver cytochrome P-450 nifedi  
A/Reference number: A32199; MUID:89093163; PMID:2463251  
A/Accession: A32199  
A/Molecule type: protein  
A/Residues: 'X', 2-9, 'XXX', 13-15 <BOR>  
R/Komori, M.; Hashizume, T.; Ohi, H.; Miura, T.; Kitada, M.; Nagashima, K.; Kamataki, T.  
J. Biochem. 104, 912-916, 1988  
A/Title: Cytochrome P-450 in human liver microsomes: high-performance liquid chromatogra  
A/Reference number: PX0012; MUID:89214010; PMID:3243766  
A/Accession: PX0012  
A/Molecule type: protein  
A/Residues: 'X', 2-11, 'X', 13-25 <KOM>  
A/Experimental source: liver microsome  
C/Genetics:  
A/Gene: GDB:CYP3A4  
A/Cross-references: GDB:118782  
A/Map position: 7q22.1-7q22.1  
C/Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C/Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
F:302-464/Domain: cytochrome P450 homology <P45>  
F:442/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 14.0%; Score 398.5; DB 2; Length 503;  
Best Local Similarity 28.7%; Pred. No. 1.5e-18;  
Matches 153; Conservative 88; Mismatches 195; Indels 97; Gaps 24;

QY 13 LAAFSW--ASIAFFSLYLAPRRS--SLYNLOG-PNHNTY-FTGNFLDILSARTG-----E 61  
DB 7 LAMETWLLAVSLVLLVLYGTHSHGLFKLGIPTPLPLGN---ILSYHKGFCMFDM 63

QY 62 BHAKYREKYGSTLRFAGIAGAPVLSNTPDKVFNHVM-KEAYD---PKPGMAARVLRIAT 117  
DB 64 CHKKYKVGWGF---YDG--QQPVLAITDPDMIKLVKVCYSVFTNRRPFGVGMKSA 117

QY 118 GGVVTAEGEAHKEHRRIMIPSLSAQAVKSMVPIFILEKGMELVDKMDAAEKDMAVGS 177  
DB 118 ---ISIADEEWKRLRSLSPFTSGKLEKMPVIAQYGVLVNLRREA----- 164

QY 178 AGEKATRILET-EGVDVKWVGVRATLDVVALAGFDYKSDSLQNKTNELYVAFLGLTDGFA 236  
Db 165 -----ETGCKFTLKDVFCAYSMDVITSTSGVNIDSLNNPD----- 201  
QY 237 PTLDSKAIM-WDFV-PYERMK--RRHEIPLTQGLAV-----SRRVGIEMEQKKQAVLGSA 290  
Db 202 PFVENTKLRLPDLDPFLLSIIVFPFLIPLEVINICVPREVTNLRKSVR-----M 256  
QY 291 SDQAVDKVOGRDILSLVRANIAANLPESQKLSDEVLAAQISNLLFAGYETSSTVLTW 350  
Db 257 KESRLEDTKHRVDFLOLMIDSQSKTESHKALSDLVLAQSIIIFAGVETTSSVLSF 316  
QY 351 MEHRLSEDAVODKLREIECOI---DTDMPTLDLINALPYLEAFVKESIRLDPSPSYANRE 408  
Db 317 IMYLATHPDVQQLQEIDEIVALEPNKAPTYDTVLQMEYLDVMVNETLRFFAMRLERV 376  
QY 409 CLKDEDPIAEPVIGRDGSVINERITKGTMVMLPLEFNINRSKFICYGEAEFRPRRWL 468  
Db 377 CKKDVE-----INGMFIPKGVWMIPSYALHRDP-KWTTSPEKFLPERFS 420  
QY 469 EDVTDLSNS-LEAPYGHQAQFISGPRACFGWRFVAVAKAFLEFVTLRRVOPBP 520  
Db 421 KKKNKDNDPIYTTPFG-----SGPRNCIGMRFAIMNMKLIALRVLFNQSFKP 467

RESULT 8  
S50211  
cytochrome P450 3A13 - mouse  
N/Contains: oxidoreductase (EC 1.-.-.-)  
C/Species: Mus musculus (house mouse)  
C/Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C/Accession: S50211; S18155  
R/Yanagimoto, T.; Itoh, S.; Sawada, M.; Hashimoto, H.; Kamataki, T.  
Biochim. Biophys. Acta 1201, 405-410, 1994  
A>Title: Molecular cloning and functional expression of a mouse cytochrome P-450 (Cyp3a-  
A/Reference number: S50211; MJUD:95101705; PMID:7803471  
A/Accession: S50211  
A/Molecule type: mRNA  
A/Residues: 1-503 <YAN>  
A/Cross-references: UNIPROT:Q64464; EMBL:X63023; NID:g50634; PIDN:CAA44754.1; PID:g50635  
C/Genetics:  
A/Gene: CYP3A13  
C/Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C/Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
F/302-464/Domain: cytochrome p450 homology <P45>  
F/442/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 14.0%; Score 398; DB 2; Length 503;  
Best Local Similarity 27.0%; Pred. No. 1.6e-18;  
Matches 142; Conservative 74; Mismatches 179; Indels 130; Gaps 20;

QY 63 HAKYREKYGTSLRFAGIAGAPLVNSTDPKVNFHM-KEAYD-----YPKPGMAARVLRL 114  
Db 65 HKYGYGMWG---LYDG-RQPVLAITDDPIIKTVLKVCYSTFTNRRRGGPVGILKKATS 119  
QY 115 IATGDGVVTAEGEAHKKRRIMTPSLSAQAVKSMVPIFEKGMELVDKXMEDAEKDMAV 174  
Db 120 I-----SENBEWKRIALLSPFTTSGLKEMFPFI-----NOFTDVLVENMRQ 162  
QY 175 GESAGEKPAKRLTEGVVDKDWVGRATLDVVALAGDYKSDSLQNKT----- 221  
Db 163 G-LGEGKPT-----SMKDI FGASMDVITATSGVNIDSLNNPDQPFVEKIKLLXF 213  
QY 222 ---NELYVAFVGLTDGAFTLDSFKAIMW--DVVPFRTRKRHEIPTLCGLAVSRVGI 276  
Db 214 DIFDPLFLS-VTLFPFLTPVFDALNVSLFPDRVISFTT-----SVERMKEN 259  
QY 277 ELMEQKKQAVLGSDAQAVDKDQVGRDILSLVRANIAANLPESOKLSDEEVLAAQISNL 336  
Db 260 RYKEKEQRV-----DFQLMINSONYTKESHKALSDVHIVAQSVIF 302  
QY 337 LFAGYTSSTVLTWPHRLISEDKAVODKLREEICQIDTMP-----TLDELNALPYLEAF 391  
:|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

Db 303 IFAGYETTSSALSFALYLLAIHPDVOKQLQDE---IDAAALPNKAPATYDTLLQMEYLDWMV 359  
 Qy 392 VKESLRDPPSPYANRECLKDEDFIPLAPVIGRDSGVINEVRIITGTVMLPLFNINRS 451  
 Db 360 VNETRLRYPFAGRLERVCKTDE-----INGLFIKPTGVMTPTFALHKD 404  
 Qy 452 KFIYGEDABEERFERMLEDTVSLNSIEAPYHQASFIISGRACFGWRFAVAMKAFLEW 511  
 Db 405 P-KWPEPEERFERRSKKNQDSIN---PYMY-LFFGSGPNCIGWRFALINMKVALVR 458  
 Qy 512 TLRAVQPEPI-----ISHPEYEHITLIISRPRIVGRE 543  
 Db 459 VLQNFVTVPCKETEIPILKLSKQGLQFENPFLLLKVVSRDETYSDE 503

RESULT 9  
 S28168  
 nifedipine oxidase (EC 1.14.14.-) cytochrome P450 3A8 - crab-eating macaque  
 N:Alternate names: cytochrome P450 CMLC; cytochrome P450 Mknf2; cytochrome P450-MK2  
 C:Species: Macaca fascicularis (crab-eating macaque)  
 C:Date: 28-May-1993 #sequence revision 28-May-1993 #text\_change 09-Jul-2004  
 C:Accession: S28168; S04509; S36875  
 R:Komori, M.; Kikuchi, O.; Sakuma, T.; Funaki, J.; Kitada, M.; Kamataki, T.  
 Biochim. Biophys. Acta 1171, 141-146, 1992  
 A:Title: Molecular cloning of monkey liver cytochrome P-450 cDNAs: similarity of the pr  
 A:Reference number: S28166; MUID:93129612; PMID:1282830  
 A:Accession: S28168  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-503 <COM>  
 A:Cross-references: UNIPROT:P33268; GB:S53047; NID:G263692; PIDN:AAB24952.1; PID:G263692  
 R:Ohta, K.; Kitada, M.; Hashizume, T.; Komori, M.; Ohi, H.; Kamataki, T.  
 Biochim. Biophys. Acta 996, 142-145, 1989  
 A:Title: Purification of cytochrome P-450 from polychlorinated biphenyl-treated crab-ea  
 A:Reference number: S04509; MUID:89287352; PMID:2500151  
 A:Accession: S04509  
 A:Molecule type: protein  
 A:Residues: 1-22 <OHT>  
 R:Ohmori, S.; Horie, T.; Guengerich, F.P.; Kiuchi, M.; Kitada, M.  
 Arch. Biochem. Biophys. 305, 405-413, 1993  
 A:Title: Purification and characterization of two forms of hepatic microsomal cytochrome  
 A:Reference number: S36874; MUID:93384294; PMID:8373178  
 A:Accession: S36875  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-13 <OHM>  
 C:Genetics:  
 A:Gene: CYP3A8  
 C:Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
 C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
 F:302-464/Domain: cytochrome P450 homology <P45>  
 F:442/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.9%; Score 395.5; DB 2; Length 503;  
 Best Local Similarity 28.6%; Pred. No. 2,4e-18;  
 Matches 155; Conservative 85; Mismatches 187; Indels 115; Gaps 26;

Qy 13 LAAPSW--ASIAFFSLYLAPRS--SLYNLQ--PNHTNY--FTGNFLDILSARTG-----E 61  
 Db 7 LAVETWLLLAATVLLVLLYLYGTSHGLFKGLGPGTPLPLLCN---ILSKRGFWTFDME 63  
 Qy 62 EIAKYREKYGSTLRFAGIAGAPVLNSTPKVFNHW--KEAYD-----YKPGKAARVL 113  
 Db 64 CYKKYGRKWGF---YDG--RQVLAITDPNMIKTVLVKECYSVFTNRPPFGVGFMKVAI 118  
 Qy 114 RIATGDGVVTABGEAHKRRIRIMPSLSAQAVKSMVPFLEKGMELVDKMDAAEKDMA 173  
 Db 119 SI-----ADEDEKWKIRSLSPTFSGKLKEMVPIIAKYGVDLVRLNREA----- 164  
 Qy 174 VCESAGEKATLET-EGYDVKDWVGRATLDYMALAGFDYKSDSLQNTKNELYVAVGLT 232  
 Db 165 -----ETGKPTVLKDFVAGSMDEVITSTSEGVNIDSLNPPQD----- 201

QY 233 DGRAPTILDSKAIM-WDFV-PYRTMK-RBHEIPLTOGLAVS---RRVGIELMEQKQAV 286  
 Db 202 -----FFVETKXLLRDFLDFPFFLSITIFFPIPIILEVLNISFPREVTSFLNKSVR1- 256  
 QY 287 LGSASDAQVKKDQVGR--DILSLVRANTAANLPESQKLSDEEVLQAIQSNLLFAGYETS 344  
 Db 257 -----KESRLKDTQKHRVDFLQLMIDSONSKETESHKALSDLELVAQSIIFIFAGYETT 310  
 QY 345 STVLTWFMHRLSEDKAVODKLRBEICQIDTDM-----PTLDELNALPYLEAFVKESRLD 399  
 Db 311 SSVLSFTIYELATHPDVQKQLOBE---IDTVLPNKAPPTDYTLQMEYLDVMVNETLRIF 367  
 QY 400 PPSYANRECLKDBDFIPLAEPVIGRDGVSINEVRITKGTVMVLPFLFNINRSKFIYGEDA 459  
 Db 368 PIANRLERVCKQV-----INGIFIPKGVVMPSPYALHHDH-KYWPPEP 411  
 QY 460 BEFRPERLWEDVDSLNS-IEAPYGHQASIFSGPRACFGWRPFAVAKAPFLFVTLRRVQF 518  
 Db 412 EKFLPERFSKQNDNIDPYIYTPFG-----SGPRNCIGNRFALMMKLAIRVLQNFSP 465  
 QY 519 EP 520  
 Db 466 KP 467

RESULT 10  
 A60564  
 Cytochrome P450 3A11 - mouse  
 N:Alternate names: cytochrome P450 3A-UT; cytochrome P450 11I1M1  
 N:Contains: oxidoreductase (EC 1.1.1.11)  
 C:Species: Mus musculus (house mouse)  
 C>Date: 17-Apr-1993 #sequence\_revision 17-May-1996 #text\_change 09-Jul-2004  
 C:Accession: S22334; A60564  
 R:Yanagimoto, T.; Itoh, S.; Muller-Enoch, D.; Kamataki, T.  
 Biochim. Biophys. Acta 1130, 329-332, 1992  
 A:Title: Mouse liver cytochrome P-450 (P-45011I1M1): its cDNA cloning and inducibility  
 A:Reference number: S22334; MUID:92223116; PMID:1339292  
 A:Accession: S22334  
 A:Molecule type: mRNA  
 A:Residues: 1-504 <YAN>  
 A:Cross-references: UNIPROT:Q64459; EMBL:X60452; NID:505034; PIDN:CAA42981.1; PID:950535  
 R:Bornheim, L.M.; Correia, M.A.  
 Mol. Pharmacol. 38, 319-326, 1990  
 A:Title: Selective inactivation of mouse liver cytochrome P-45011I1A by cannabidiol.  
 A:Reference number: A60564; MUID:90384441; PMID:2402224  
 A:Accession: A60564  
 A:Molecule type: protein  
 A:Residues: 1-17, 'X', 19-20, 'X', 22-24 <BOR>  
 A:Comment: This member of the cytochrome P45011I1A family was designated UT because it ca-  
 lation of testosterone.  
 C:Genetics:  
 A:Gene: Cyp3a11  
 C:Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
 C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
 F:303-465/Domain: cytochrome P450 homology <P45>  
 F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.8%; Score 394; DB 2; Length 504;  
 Best Local Similarity 25.6%; Pred. No. 3e-18;  
 Matches 144; Conservative 103; Mismatches 215; Indels 100; Gaps 23;

QY 7 LTGALGLAAGWASIAFFSLY-LAPRRSSLYNLOQ-----PNHTNYFTGNFLDILSAR 58  
 Db 3 LVASLSLTWVLLAISLVLLRYGTRKHELFKKQIGPKPLP-----FLGTVLNYYKGL 57  
 QY 59 TGEHAKYREKYGSTL-RFAGIAGAPVLNSTDPKVFNHVM-KEAY-----DYPKQMA 109  
 Db 58 WKFDMECYK-KYGTGWLFDG--QTPLAVIDPTIKNLVYKCFSVFTNRDRFGPVGIM 114  
 QY 110 ARVLRIATGDGVWTAEGEAKHRRIRIMIPSLSAQAKSMVPIFLEKGMELVDKQMEDAAE 169  
 Db 115 SKAISISKDD-----ENKRYALLSPTTSGLKEMFFVIEQYG-DILVKYLRQKAK 165

QY 170 KDMAVGESAGEKATRLTEGVVDKQWVGRATLDVWALAGFDYKSDSLQNKTNELYVAFV 229  
 Db 166 KG-----KPVTKMDVLGAYSDVITSTSGVNVDSLNNPDPFVEKAK 208  
 QY 230 GLT--DGRAPTILDSFKAIMDWFV-PYRTMKRRIHEIPLTOGLAVSRVRVGTIELMEQKQAV 286  
 Db 209 KLLRDFDFDPL--FSVVLFPFLTPVYEML-----NICMPKDSIEFFK--F 252  
 QY 287 LGSASDAQVKKDQVGRDILSLVRA-NIAANLPESQKLSDEEVLQAIQSNLLFAGYETS 345  
 Db 253 VDRKSEKLSQKQHRVDFLQLMVNSHNKSKVSHKALSDMEITAQSIIFIFAGYETS 312  
 QY 346 TVLTWFMHRLSEDKAVODKLRBEICQI--DTDMTLDELNALPYLEAFVKESRLDPPSP 403  
 Db 313 STLSTLHSLATHPDIOKKLODEIDALPNKAPPTDYTMEMEYLDVMVNETLRLYPIAN 372  
 QY 404 YANRECLKDBDFIPLAEPVIGRDGVSINEVRITKGTVMVLPFLFNINRSKFIYGEDAEFR 463  
 Db 373 RLERVCKQV-----LNGVYIPKGVVMPSPYALHHDH-PHFQ 416  
 QY 464 PERMLEDYDLSNS-IEAPYGHQASIFSGPRACFGWRPFAVAKAPFLFVTLRRVQFPII 522  
 Db 417 PERTSKENKGSIDPVYLPFG-----NGPRNCIGMRFALMMKLAITKIMQNFSPQFC- 469  
 QY 523 SHPEYEHITLISRPVIGREK 544  
 Db 470 ---KETQIPKLKSLQGLLQPEK 488

RESULT 11  
 JX0062  
 Cytochrome P450 3A7 - human  
 N:Alternate names: cytochrome P450 3A3; cytochrome P450 HFL33; cytochrome P450 HLP2; cyt  
 N:Contains: oxidoreductase (EC 1.1.1.11)  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
 C:Accession: JX0062; PX0014; S04983; S02152  
 R:Komori, M.; Nishio, K.; Oh, H.; Kitada, M.; Kamataki, T.  
 J. Biochem. 105, 161-163, 1989  
 A:Title: Molecular cloning and sequence analysis of cDNA containing the entire coding res  
 A:Reference number: JX0062; MUID:89255154; PMID:2722762  
 A:Accession: JX0062  
 A:Molecule type: mRNA  
 A:Residues: 1-503 <KOM>  
 A:Cross-references: UNIPROT:P24462; GB:D00408; NID:9220148; PIDN:BAA00310.1; PID:9220149  
 A:Experimental source: fetal liver  
 R:Komori, M.; Hashizume, T.; Oh, H.; Miura, T.; Kitada, M.; Nagashima, K.; Kamataki, T.  
 J. Biochem. 104, 912-916, 1988  
 A:Title: Cytochrome P-450 in human liver microsomes: high-performance liquid chromatogra  
 A:Reference number: PX0012; MUID:89214010; PMID:3243766  
 A:Accession: PX0014  
 A:Molecule type: protein  
 A:Residues: 1-25 <KO2>  
 A:Experimental source: liver microsome  
 R:Komori, M.; Nishio, K.; Fujitani, T.; Oh, H.; Kitada, M.; Mima, S.; Itahashi, K.; Kam  
 Arch. Biochem. Biophys. 272, 219-225, 1989  
 A:Title: Isolation of a new human fetal liver cytochrome P450 cDNA clone: evidence for e  
 A:Reference number: S04983; MUID:89286124; PMID:2786707  
 A:Accession: S04983  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 282-503 <KO3>  
 R:Wrighton, S.A.; Vandenbranden, M.  
 Arch. Biochem. Biophys. 268, 144-151, 1989  
 A:Title: Isolation and characterization of human fetal liver cytochrome P450H2: a thir  
 A:Reference number: S02152; MUID:89104413; PMID:2492179  
 A:Accession: S02152  
 A:Molecule type: protein  
 A:Residues: 1-11, 'X', 13-25, 'X', 27-30 <WRI>  
 C:Genetics:  
 A:Gene: GDB:CYP3A7  
 A:Cross-references: GDB:134409  
 A:Map position: 7q22.1-7q22.1

C;Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
F;302-464/domain: cytochrome P450 homology (P450)  
F;442/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.5%; Score 385.5; DB 2; Length 503;  
Best Local Similarity 29.0%; Pred. No. 1.1e-17;  
Matches 157; Conservative 79; Mismatches 191; Indels 115; Gaps 26;

QY 13 LAAPSW--ASTAFFSLYL-APRRSLYNLOG-PNHTNY-FTGNFLDILSARTG-----E 61  
DB 7 LAVETWLLAVSLVLLYLGTRTHGLFKRLGIPGPTPLPLGN---ALSFRKGWTFDWE 63

QY 62 EHAKYREKYGSTLRFAGTAGAPVLNSTDPKVFNVHM-KEAYD-----YPKQGMARVL 113  
DB 64 CYKYRKVWG-----IVDCQPMALAITDPMKIVLVKECYSVFTNRPSPGPGVGMKNAI 118

QY 114 RIATGDDGVVTAEGEAHRRHIMIPSLSAQAVKSMVPIFEKGMELVDKMDAEEKDMA 173  
DB 119 SI-----AEDEWKRIIRSLSPFTTSGKLKEMVPIIAQYGDVLVRLRREA----- 164

QY 174 VBSAGEKKATRLT-EGVDVKWVGRTALDVMALAGFDYKSDSLQNKTNELVAVFGLT 232  
DB 165 -----ETGKPVTLKHVFGAYSDVITSTSGVSIIDSLNPNQD-----PFFVNT 207

QY 233 D---GFAPTLDLDFKAIMWDFVPVPRTRMKRRHEIPLTQGLAVS---RRVGIELMEQKKQAV 286  
DB 208 KKLIRFNP-LDPFVLSIKVP-PFL-----TPILEALNITVPKRVISFLSKVKQIK 257

QY 287 LGSASDAQVDKQVQGR--DILSLVRANITANLPSQKLSDEEVLQAISNLLPAGVETS 344  
DB 258 EGRU-----KETQKHURVDFLQIMDSQNSKSETHKALSDLELMAQSIIIFPAGVET 310

QY 345 STVLTWFMHRLSEDKAVQDKLREEICQIDTDM-----PTLDELNALPYLEAFVKESLRLD 399  
DB 311 SSVLSFIYELATHPDVQVKQKE---IDTVLPNKAPPTVDTVLQLEYLDMVNVNTRLRF 367

QY 400 PPSPYANRECLKDEDFPLAEPVIGRDSVINEVRITKGMVLMPLFNINRSKFIYGEDA 459  
DB 368 PVAMRLSERVCKQVE-----INGMFIPKGVVVMIPSVLVLHHP- KYWTEP 411

QY 460 EEPERPERWLEDVTDLSNS-IEAPYHQASFISCPACFGWRFAVEMKAFFLTLLRRVQF 518  
DB 412 EKFLPERFSKQKVNIDPIYTPFG-----SGPRNCIGRFPALVMKIALVQLVQNFSP 465

QY 519 EP 520  
DB 466 KP 467

RESULT 12  
A34101  
cytochrome P450 3A5 - human  
N;Alternate names: cytochrome P450 H1p2  
N;Contains: oxidoreductase (EC 1.-.-.)  
C;Species: Homo sapiens (man)  
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
R;Accession: A34101; S06491; I52302  
R;Aoyama, T.; Yamano, S.; Waxman, D.J.; Lapenson, D.P.; Meyer, U.A.; Fischer, V.; Tyndal  
J. Biol. Chem. 264, 10388-10395, 1989  
A;Title: Cytochrome P-450 hPCN3, a novel cytochrome P-450 IIAA gene product that is diff  
DNA-expressed hPCN1 and hPCN3 for the metabolism of steroid hormones and cyclosporine.  
A;Reference number: A34101; MUID:89278095; PMID:2732228  
A;Accession: A34101  
A;Molecule type: mRNA  
A;Residues: 1-502 <AOY>  
A;Cross-references: UNIPROT:P20815; GB:J04813; NID:G181345; PID:AAA02993.1; PID:G181346  
R;Schuetz, J.D.; Molowa, D.T.; Guzelian, P.S.  
Arch. Biochem. Biophys. 274, 355-365, 1989  
A;Title: Characterization of a cDNA encoding a new member of the glucocorticoid-respons  
A;Reference number: S06491; MUID:90025114; PMID:2802615  
A;Accession: S06491  
A;Status: not compared with conceptual translation

A;Molecule type: mRNA  
A;Residues: 1-304, P', 306-317, P', 319-323, D', 325-376, G', 378-502 <SCH>  
R;Jounaidi, Y.; Guzelian, P.S.; Maurel, P.; Villare, M.S.  
Biochem. Biophys. Res. Commun. 205, 1741-1747, 1994  
A;Title: Sequence of the 5'-flanking region of CYP3A5: comparative analysis with CYP3A4  
A;Reference number: I52302; MUID:95110318; PMID:7811260  
A;Accession: I52302  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-24 <RES>  
A;Cross-references: GB:S74699; NID:G786472; PIDN:AAD14157.1; PID:G4261857  
C;Genetics:  
A;Gene: GDB:CYP3A5  
A;Cross-references: GDB:118783  
A;Map position: 7q22.1-7q22.1  
C;Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
F;302-463/domain: cytochrome P450 homology (P450)  
F;441/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.5%; Score 385; DB 1; Length 502;  
Best Local Similarity 28.1%; Pred. No. 1.2e-17;  
Matches 151; Conservative 84; Mismatches 196; Indels 106; Gaps 25;

QY 13 LAAPSW--ASTAFFSLYL-APRRSLYNLOG-PNHTNY-FTGNFLDILSARTG-----E 61  
DB 7 LAVETWLLAVSLVLLYLGTRTHGLFKRLGIPGPTPLPLGN---VLSYRQGLWKFDTE 63

QY 62 EHAKYREKYGSTLRFAGTAGAPVLNSTDPKVFNVHM-KEAYDYPKPGMAARVLRTATG-- 118  
DB 64 CYKYRKVWG---YEG--QLPVLAITDPVIRTVLVEKY-----SVFNRRSLGPV 111

QY 119 -----DGVVTAEGEAHRRHIMIPSLSAQAVKSMVPIFEKGMELVDKMDAEEKDMAV 174  
DB 112 GPMKSAISLAEDDEWKRIIRSLSPFTTSGKLKEMFPIIAQYGDVLVRLRRE-AEKG-- 167

QY 175 GESAGEKKATRLTETEGVDVKWVGRTALDVMALAGFDYKSDSLQ-----KTNELYVAF 228  
DB 168 -----KPVTLKDIFGAYSDVITSTSGVSIIDSLNPNQDFFVESTKKFKF 213

QY 229 VGLTDGFAPTLDLDFKAIMWDFVPVPRTRMKRRHEIPLTQGLAVS---RRVGIELMEQKKQAV 286  
DB 214 -----GFLDPL-FLSILFPFL-----TPVPEALNVSLFPKDTINFLSKS---- 252

QY 287 LGSASDAQVDKQVQGRDILSLVRANITANLPSQKLSDEEVLQAISNLLPAGVETSST 346  
DB 253 VNRMEKSLRNDKQKRLDFLQIMDSQNSKSETHKALSDLELMAQSIIIFPAGVETSS 312

QY 347 VLTWFMHRLSEDKAVQDKLREEICQI--DTDMPTLDELNALPYLEAFVKESLRLDPPSPY 404  
DB 313 VLSFTLYELATHPDVQVKQKEIDAVLPNKAPPTVDAVQVMEYLDVMVNETIRLFPVAIR 372

QY 405 ANRECLKDEDFPLAEPVIGRDSVINEVRITKGMVLMPLFNINRSKFIYGEDAEFRP 464  
DB 373 LERTCKQVE-----INGVFIPKGSVVVPIYALHHPD-KYWTSPPEFRP 416

QY 465 ERMLEDVTDLSNS-IEAPYHQASFISCPACFGWRFAVEMKAFFLTLLRRVQFEP 520  
DB 417 ERFKSK-KDSIDPIYTPFG-----TGRNCIGRFPALVMKIALIRVLQNFSEKFP 466

RESULT 13  
A29487  
cytochrome P450 3A6 (version 1) - rabbit  
N;Alternate names: cytochrome P450 3C  
N;Contains: oxidoreductase (EC 1.-.-.)  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 09-Jul-2004  
C;Accession: A29487  
R;Dale, C.; Clair, P.; Daujat, M.; Fort, P.; Blanchard, J.M.; Maurel, P.  
DNA 7, 39-46, 1988  
A;Title: Complete sequence of cytochrome P450 3c cDNA and presence of two mRNA species  
A;Reference number: A29487; MUID:88166352; PMID:3349903

A:Accession: A29487  
A:Molecule type: mRNA  
A:Residues: 1-501 <DAL>  
A:Cross-references: UNIPROT:P11707; GB:M19139; NID:G165573; PIDN:AAA31430.1; PID:G165574  
C:Genetics:  
C:Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:301-462/Domain: Cytochrome P450 homology <P45>  
F:440/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.3%; Score 379; DB 2; Length 501;  
Best Local Similarity 26.2%; Pred. No. 2.8e-17;  
Matches 147; Conservative 91; Mismatches 208; Indels 114; Gaps 22;

QY 19 ASIAPFSLYAPRSSLYNQGNHTNY-FTGNFLDILSARTGEEHAKYREKYGSTLREA 77  
Db 15 ASVLVLYGTSHTGLPKKMGICGPTPLPIGILILEVRKVINWDFIECRKKYG---KWM 71  
QY 78 GI--AGAPVLNSTDPKVFNVHM-KEAY-----DYPKPGMAARVLRIRATGCGVVTAGE 127  
Db 72 GLFDGROPLAVITDPDMDIKTVLKECVSVFTNRRSGFVGFMKAVSI-----SEDE 123  
QY 128 AHKRRHRIIMPISAAQAVKSMVPFLEKGMELVDKMDAAEKDMDVAVGESAGEKATRL 187  
Db 124 DWKVRILLPTFTSGKLEMLPIIAQYGDVLVKNLROEAEKASPST----- 170  
QY 188 TEGVDVDMVGRATLDVVALAGFDYKSDSLQNKNE-----LYVAFVGLDGFAPTL 239  
Db 171 ----LKEIFGAYGMDVITGTSFGWIDSLRNQDPFKNVRLLKFSF-----FDPLL 219  
QY 240 DSFAIMWDFVPYFRTMKRRHEPLTQGLAVSRVGVLEMEQKQAVLGASDAQVDKXD 299  
Db 220 LSI-----TUPPFL-----TFIFALHIS-MFKDVMDFKTSVEKIKDRLDKQK 265  
QY 300 VQGRDILSLVRANIAANLPESQKLSDEEVLQIISNLLFAGYETSTSLVTFMFRHLSDEK 359  
Db 266 -RRVDFQLMINSQNSKXIDSHKALDIEVVAQSIILFAGYETSTSLFIMHLLATHP 324  
QY 360 AVQKLEBEICQITDNP-----TLDELNALPYLEAFVKESILDPSPVYANRECLKDE 414  
Db 325 DVQOKLOE-----IDTLLENKELATYDLVKVEYLDVVMVNETLRYTAGLERVCKVD 381  
QY 415 FIPLAEPVIGRDSVINEVITKTGMVWMLPFLNINRSKFTYGDABEFPERWLEVDVDS 474  
Db 382 -----INGTIPKGTIVM-PTYALHRDPQHWTE-PDEFPERFSKKNKN 424  
QY 475 LNS-IEAPYHQASFIQPRACFGWRFAVEMKAFELVTLRRVQFE-----PI----- 521  
Db 425 INPIYHPFG-----AGPRNCLGMRPALNMIKALVRLMQNFSFKLCKETQVPLKLGQ 478  
QY 522 -ISHPEVEHITLISRPRIV 540  
Db 479 GLLOPERPIVLKVVSESGII 498

RESULT 14  
A22631  
Cytochrome P450 3A1, pregnenolone 16-alpha-carbonitrile-Inducible - rat  
N:Alternate names: testosterone 6beta-hydroxylase  
N:Contains: unspecific monooxygenase (EC 1.14.14.1) cytochrome P450 PCN1  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 09-Jul-2004  
C:Accession: A22631; PX0035; S21697; S36137; S27107; S30378; I59218  
R:Gonzalez, F.J.; Nebert, D.W.; Hardwick, J.P.; Kasper, C.B.  
J. Biol. Chem. 260, 7435-7441, 1985  
A:Title: Complete cDNA and protein sequence of a pregnenolone 16-alpha-carbonitrile-inducible cytochrome P450 3A1  
A:Reference number: A22631; MUID:85207783; PMID:3838989  
A:Molecule type: mRNA  
A:Residues: 1-504 <GN>  
A:Cross-references: UNIPROT:P04800; GB:M10161; NID:G203777; PIDN:AAA41035.1; PID:G203778  
R:Nagata, K.; Gonzalez, F.J.; Yamazoe, Y.; Kato, R.

J. Biochem. 107, 718-725, 1990  
A:Title: Purification and characterization of four catalytically active testosterone 6beta-hydroxylase related forms.  
A:Reference number: PX0032; MUID:90375438; PMID:2398038  
A:Accession: PX0035  
A:Molecule type: protein  
A:Residues: 1-26 <NAG>  
A:Experimental source: liver, Sprague-Dawley male rat  
R:Lechner, M.C.  
Submitted to the EMBL Data Library, December 1991  
A:Reference number: S21697  
A:Accession: S21697  
A:Molecule type: mRNA  
A:Residues: 1-206, 'A', 208-212, 'I', 214-231, 'V', 233-504 <LEC>  
A:Cross-references: EMBL:X64401; NID:G56038; PIDN:CAA45743.1; PID:G56039  
R:Ribeiro, V.; Lechner, M.C.  
Arch. Biochem. Biophys. 293, 147-152, 1992  
A:Title: Cloning and characterization of a novel CYP3A1 allelic variant: Analysis of CYP3A1  
A:Reference number: S36137; MUID:92117688; PMID:1731631  
A:Accession: S36137  
A:Molecule type: mRNA  
A:Residues: 205-206, 'A', 208-212, 'I', 214-231, 'V', 233-234 <RIB>  
A:Cross-references: EMBL:X64401  
R:Teihada, M.B.; Pereira, T.M.; Lechner, M.C.  
Arch. Biochem. Biophys. 298, 715-725, 1992  
A:Title: Effect of dexamethasone and phenobarbital on run-on transcription rate and CYP3A1  
A:Reference number: S27107; MUID:93037516; PMID:1417000  
A:Accession: S27107  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-24 <TEL>  
A:Cross-references: EMBL:X62086  
R:Cooper, K.O.; Reik, L.M.; Jayyosi, Z.; Bandiera, S.; Kelley, M.; Ryan, D.E.; Daniel, R.  
Arch. Biochem. Biophys. 301, 345-354, 1993  
A:Title: Regulation of two members of the steroid-inducible cytochrome P450 subfamily (3A1 and 3A2) by dexamethasone and phenobarbital  
A:Reference number: S30378; MUID:9213168; PMID:7681660  
A:Accession: S30378  
A:Molecule type: protein  
A:Residues: 1-25 <COO>  
R:Burger, H.  
Proc. Natl. Acad. Sci. U.S.A. 89, 2145-2149, 1992  
A:Title: Paradoxical transcriptional activation of rat liver cytochrome P-450 3A1 by dexamethasone in primary monolayer cultures of adult rat hepatocytes.  
A:Reference number: I59218; MUID:92196074; PMID:1372436  
A:Accession: I59218  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-24 <BUR>  
A:Cross-references: GB:M86850; NID:G205919; PIDN:AAA41780.1; PID:G205920  
C:Genetics:  
C:Gene: CYP3A1; P450p  
C:Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:303-465/Domain: Cytochrome P450 homology <P45>  
F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.0%; Score 370; DB 2; Length 504;  
Best Local Similarity 26.2%; Pred. No. 1.1e-16;  
Matches 147; Conservative 84; Mismatches 189; Indels 140; Gaps 23;

QY 7 LTGALGLAAPSASIAFSLY-LAPRSSLYNLOG-----PNHTNYFTGNF-LDI 54  
Db 3 LLSAULTWTWLLAVLVLLVGLFTRHGLFKKGIQPKPLPFCFVLYNGLWKFV 62  
QY 55 LSARTGEEHAKYREKYGSTLRFAGIAGAPVLNSTDPKVFNVHM-KEAY-----DYPKP 106  
Db 63 -----ECHKYKGIWG---LFDG-QMPLFAITDTEMKNLVKECFSVFTNRDFGPV 111  
QY 107 GMAARVLRIRATGCGVVTAGEAHKRRHRIIMPISAAQAVKSMVPFLEKGMELVDKMD 166  
Db 112 GI-----NGKAVSNAKSEWKRKRALISPTTSRLEKEMFFIIQYGDILVKYKQE 163  
QY 167 AAEXDMDVAVGESAGEKATRLLET-EGVDYKDWVGRATLDVVALAGFDYKSDSLQ 219



Db 164 A-----ETGKPVTKKVFSGAYSDVITSTSGVNVDSLNPKDPFV 204  
Qy 220 -KTNEL-----YVAFVGLTQGFAPTLDSFKAIMW--DFVPYFRMKRRHEIPL 264  
Db 205 EKTKLLRFDFFDLFLSVLFPPLT-----PIYEWLNICMPFKDSIEFFK----- 250  
Qy 265 TQGLAVSRVGIEMEQKQAVLGASDAQDKQVQGRDILSLVRANIAANLPESQ-K 323  
Db 251 -----KPYRKETRLDSVQKHRV-----DFLQMMNAHNSDKDKESHYA 290  
Qy 324 LSDEEVLAQISNLLFAGYETSTVLTWTFHRLSBDKAVQDKLREEICQI--DTDMPTLDE 381  
Db 291 LSDMEITAQSIIFIFAGYEPTSTLSFVLHSLATHPTQKQLQEEIDBALPNKAPPTYDT 350  
Qy 382 LNALPYLEAFVKESLRLDPPSPYANRECLDKEDFIPLAEPVIGRDGVSINEVRITKTMV 441  
Db 351 VMEMEYLDMLNLTRLPYIGNRLRVCKKQVE-----INGVFWPKGSVV 395  
Qy 442 MLPLFNINRSKFYIGEDAEERPRWLEDVTDLSNS-TEAPYGHQASFISGPRACFGWR 500  
Db 396 MIPSYALHRDPQHWP-PEEFPFRFSKNGKSIDPVYLPFG-----NGPRNCIGMR 448  
Qy 501 AVAEMKAPFLVTLRRVOFEP 520  
Db 449 ALNMKALATKVLQNFSPQ 468

RESULT 15  
S50892  
Cytochrome P450 3A16 - mouse  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mus musculus (house mouse)  
C>Date: 13-Jan-1995 #sequence\_revision 09-May-1997 #text\_change 09-Jul-2004  
C:Accession: S50892  
R:Itch, S.; Satoh, M.; Abe, Y.; Hashimoto, H.; Yanagimoto, T.; Kamataki, T.  
Eur. J. Biochem. 226, 877-882, 1994  
A:Title: A novel form of mouse cytochrome P450 3A (Cyp3a-16). Its cDNA cloning and expression  
A:Reference number: S50892; MUID:95112853; PMID:7813478  
A:Accession: S50892  
A:Molecule type: mRNA  
A:Residues: 1-504 <ITO>  
A:Cross-references: UNIPROT:Q64481; EMBL:D26137; NID:g493670; PIDN:BAAC05133.1; PID:g6669  
A:Experimental source: fetal liver  
A:Note: in the authors' translation residues 9-20 do not match the nucleotide sequence  
A:Note: the authors translated the codon TTC for residue 464 as Glu  
C:Genetics:  
A:Gene: CYP3A-16  
C:Superfamily: human cytochrome P450 CYP3A5; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F/303-465/Domain: cytochrome P450 homology <P45>  
F/469/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 12.9%; Score 368; DB 2; Length 504;  
Best Local Similarity 25.8%; Pred. No. 1.5e-16;  
Matches 132; Conservative 100; Mismatches 171; Indels 108; Gaps 21;

Qy 68 EKYGSTL-RFAGTAGAPVLNSTDPKVFNVHM-KEAY-----DYPKPGMAARVLRATG 118  
Db 66 EKYGKTWGLFDG--QIPLFVITDPTIKVNLVRECEFSFTNRQDFPFGVIMSKISL--- 120  
Qy 119 DGVVTAGEAHKRRHRRMIPSLSAQAVKSMVPIFLEKGMELVDKMDAEDAKDMAYGES 178  
Db 121 -----AKDEWKYRALLSFTTSGNLKEMFPFVIEQIG-DILVKYLRQAEKGPVA--- 171  
Qy 179 GEXKATLETEGVVDKDWGRATLDVNALAGFDYKSDSLQNKTNELNYAVFVGLTDGFAPT 238  
Db 172 -----VKDVLGAYSMDVIISTTFGVNIDSLNRPD-----PF 203  
Qy 239 LDSFKAIM-WDF-----VPYFRMKRRHEIPLTQGLAVSRVGIEMEQKQAVLGSA 290  
Db 204 VENAKVLRDYPFDPLSLVALFPFLTPYEM---LNICVFKDSIEFFK-----FVDRM 256

Qy 291 SDQAVKKDVQGRDILSLVRANIAANLPESQK-LSDEEVLAQISNLLFAGYETSTVLT 349  
Db 257 TENRLDSKQKHRVDFYLMWEAYNKSQKDSHKALSEIEITTAQSIIFIFAGYETSSILS 316  
Qy 350 WMFHLRSEDKAVQDKLREEICQI--DTDMPTLDELNALPYLEAFVKESLRLDPPSPYANR 407  
Db 317 FTVISLATHPTDIQKKLQEEIDEALPNKAPPTYDTVMAMEYLDMLNLTRLPYITNRLQR 376  
Qy 408 ECLKQDSDFIPLAEPVIGRDGVSINEVRITKTMVMLPLFNINRSKFYIGEDAEERPRW 467  
Db 377 VCKKQVE-----INGIYIPKGSVTIIPSVVLHHDHDPQHWP-PEEQPERF 420  
Qy 468 LEDVTDLSNS-TEAPYGHQASFISGPRACFGWRFAVAEMKAPFLVTLRRVOFEPISHP 526  
Db 421 SKENKSIDPVYLPFG-----NGPRNCIGMRFALNMKALIKVLQNFSPQPC---K 470  
Qy 527 YEHITLIIISRPRIVGRKEGYQMRQ-VKPV 556  
Db 471 ETQIPKLRS-----ELLQPVKPI 490

Search completed: December 14, 2004, 21:43:19  
Job time : 47 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2004, 15:23:49 ; Search time 221 Seconds  
(without alignments)  
1450.152 Million cell updates/sec

Title: US-10-066-007A-1

Perfect score: 2852

Sequence: 1 MFILVLLTGALGAFAFWAS.....RIVGREKEGYCMRLQVKPE 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425	14.9	576	2 Q8J2Q5	Q8J2Q5 gibberella
2	422.5	14.8	503	2 Q6GQ03	Q6GQ03 macaca mula
3	419	14.7	593	2 Q9C6S0	Q9C6S0 arabidopsis
4	419	14.7	595	2 Q93VK5	Q93VK5 arabidopsis
5	417	14.6	503	2 Q97689	Q97689 sus scrofa
6	417	14.6	517	2 Q8LIR5	Q8LIR5 oryza sativ
7	414	14.5	503	1 CP3T_PIG	P79401 sus scrofa
8	413	14.5	503	1 CP39_RAT	P51538 rattus norv
9	408.5	14.3	503	1 CP3C_SHEEP	Q29496 ovis aries
10	405	14.2	514	2 Q6T485	Q6T485 triticum ae
11	405	14.2	514	2 Q6DV71	Q6DV71 triticum ae
12	405	14.2	514	2 AAR11387	AAR11387 triticum
13	404	14.2	503	1 CP3C_CANFA	P24463 canis famil
14	403.5	14.1	503	2 Q95L35	Q95L35 sus scrofa
15	403	14.1	515	2 Q9LUD2	Q9LUD2 arabidopsis
16	402	14.1	501	1 CP35_RABIT	P11707 oryctolagus
17	401	14.1	535	2 Q8WLE1	Q8WLE1 arabidopsis
18	400.5	14.0	503	2 Q6PSM4	Q6PSM4 macaca mula
19	400.5	14.0	503	2 AAS91645	AAS91645 macaca mula
20	400.5	14.0	519	2 Q9ASR3	Q9ASR3 arabidopsis
21	399	14.0	503	1 CP33_HUMAN	P05184 homo sapien
22	398.5	14.0	503	2 AAS95418	AAS95418 homo sapi
23	398	14.0	503	1 CP3D_MOUSE	Q64464 mus musculu
24	397.5	13.9	519	2 Q8LHV0	Q8LHV0 oryza sativ
25	396.5	13.9	502	1 CP34_HUMAN	P08684 homo sapien
26	395.5	13.9	503	1 CP38_MACFA	P33268 macaca fasc
27	395.5	13.9	503	2 Q6VP01	Q6VP01 macaca mula
28	395.5	13.9	503	2 AAP94642	AAP94642 macaca mu
29	394	13.8	503	2 Q86SK3	Q86SK3 homo sapien
30	394	13.8	504	1 CP35_MOUSE	Q64459 mus musculu
31	393.5	13.8	544	2 Q9FEE1	Q9FEE1 oryza sativ

32 392.5 13.8 503 2 Q76N65 Q76N65 sus scrofa  
33 392.5 13.8 503 2 BAD06180 BAD06180 sus scrofa  
34 392 13.7 497 2 Q8CJF2 Q8CJF2 rattus norv  
35 391 13.7 504 1 C341\_MOUSE C341\_MOUSE mus musculu  
36 391 13.7 561 2 Q7S0S8 Q7S0S8 neurospora  
37 389 13.6 492 2 Q8ISJ7 Q8ISJ7 manestra br  
38 387 13.6 523 2 Q8LQ77 Q8LQ77 oryza sativ  
39 386.5 13.6 511 2 Q8HY76 Q8HY76 capra hircu  
40 385.5 13.5 503 1 CP37\_HUMAN CP37\_HUMAN  
41 385.5 13.5 503 2 AAS67436 AAS67436 homo sapi  
42 385.5 13.5 535 2 Q9H241 Q9H241 homo sapien  
43 385 13.5 502 1 CP35\_HUMAN CP35\_HUMAN  
44 385 13.5 502 2 AAS02016 AAS02016 homo sapi  
45 385 13.5 507 1 CP3S\_BOVIN CP3S\_BOVIN

#### ALIGNMENTS

#### RESULT 1

Q8J2Q5 PRELIMINARY; PRT; 576 AA.  
AC Q8J2Q5;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DE Fum15p.  
GN Name=FUM15;  
OS Gibberella moniliformis.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
ON NCBI\_TaxID=117187;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-3125;  
RX MEDLINE=99343882; PubMed=104113619;  
RA Proctor R.H., Desjardins A.E., Plattner R.D., Hohn T.M.;  
RT "A polyketide synthase gene required for biosynthesis of fumonisin  
mycotoxins in Gibberella fujikuroi mating population A.";  
RL Fungal Genet. Biol. 27:100-112(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-3125;  
RX MEDLINE=21585628; PubMed=11728154;  
RA Seo J.A., Proctor R.H., Plattner R.D.;  
RT "Characterization of four clustered and coregulated genes associated  
with fumonisin biosynthesis in Fusarium verticillioides.";  
RL Fungal Genet. Biol. 34:155-165(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-3125;  
RX MEDLINE=22508125; PubMed=12620260;  
RA Proctor R.H., Brown D.W., Plattner R.D., Desjardins A.E.;  
RT "Co-expression of 15 contiguous genes delineates a fumonisin  
biosynthetic gene cluster in Gibberella moniliformis.";  
RL Fungal Genet. Biol. 38:237-249(2003).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-3125;  
RX MEDLINE=22309955; PubMed=12423021;  
RA Desjardins A.E., Munkvold G.P., Plattner R.D., Proctor R.H.;  
RT "FUM1-a gene required for fumonisin biosynthesis but not for maize  
ear rot and ear infection by Gibberella moniliformis in field tests.";  
RL Mol. Plant Microbe Interact. 15:1157-1164(2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-3125;  
RA Butchko R.A.E., Plattner R.E., Proctor R.H.;  
RT "FUM9 is required for C-5 hydroxylation of fumonisins and complements  
the metabolically defined fum3 locus in Gibberella moniliformis.";  
RL Appl. Environ. Microbiol. 0:0-0(2003).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
EMBL; AF155773; AAN74818.1; -.

[illegible][illegible]

RA	Carroll B.A., Fang Y., Rushmore T.H.;
RL	Submitted (MAY-2004) to the ENBL/GenBank/DBDJ databases.
CC	-/- SIMILARITY: Belongs to the cytochrome P450 family.
DR	EML; AV635466; AAT49270.1; -
DR	InterPro; IPR001128; Cytochrome_P450.
DR	InterPro; IPR002401; EP450I.
DR	InterPro; IPR008072; EP450_CYP3A.
DR	Pfam; PF00067; p450; 1.
DR	PRINTS; PR00463; EP450I.
DR	PRINTS; PR01689; EP450IIICYP3A.
DR	PRINTS; PR00385; P450.
DR	PROSITE; PS00286; CYTOCHROME P450; UNKNOWN_1.
KW	Heme; Monooxygenase; Oxidoreductase.
SQ	SEQUENCE 503 AA; 57382 MW; D0B8902ADDFF33810 CRC64;
	Query Match 14.8%; Score 422.5; DB 2; Length 503;
	Best Local Similarity 29.1%; Pred. No. 8.3e+20;
	Matches 154; Conservative 85; Mismatches 201; Indels 89; Gaps 233
Qy .	13 LAAPSW--ASTAFSLYLAPRS-SLYNLOG-PNHNY-PTGNFLDILSARG--EEHAK 65
Db	7 LANEYTLALLVLLVLLDYGRSYGLFRGIGPPTLPFLNG---ILSYRGLMKFDE 63
Qy	66 YREKYGSTLRFAGIAGAPVLNSDPKVFNHM-KEAYDY---PKPGMAARVLRIATGDGV 121
Db	64 CVYKGKMWRTOD-GQLPVLTITDPEMIKTVLVKECYSVFTRRRPLGPVGLMKS- ---I 118
Qy	122 VTAEGEAHKRIRIMIISLSAQAVKSMVIPLEKGMELVDKOMEDAAEKDMAVGESAGEK 181
Db	119 STADEEWKRISSLSPFTTSGKLKENFPPIAGYGNLVNRURR--AEKG----- 167
Qy	182 KATRLETEGGVDVQWVGRAITLDVMALAGFDYKSDLSQNKTNELYAVFGLTGDFAPTLD 241
Db	168 -----KPVTLKIDFGAYSNDVITSFGVNIDSLSNP-----KDPFVESVK 209
Qy	242 FKAIMWDVP--YPRTWKRHEPLTIOGLAVS---RRVGIEMEQKQAVLGSASQAVDK 297
Db	210 F--LKFDLDFLTLTLPPFLPAFEALNVSLFPKDAINFLNKs-----VNSMKKSRLD 263
Qy	298 KDVGQRDILSLVRANTAANLPSOKLSDBEVLQAQISNLIFAGYETSSVLTMTMPHLRSE 357
Db	264 KQKHVEDFLQMDSQNSKETESHKALSDOELVAQSIIIPAGYETTSSVLSFIYELAT 323
Qy	358 DKAVODKUREBIQCIDTMP-----TLDELNALPYLEAFVKESLRLDPPSPVANRECUKD 412
Db	324 HPDVQQKIQKE---IDAVLENKAPATYDAVMQVEYLDVMVNETLRFPIATRLERACKD 380
Qy	413 EDFIPLABEVIGROGVSINVRITKGMVMVLPLNFNINRSKFYIGDADFEPFERLVEDVT 472
Db	381 VE-----INGVTFPKGMVVPIYIALHDDP-KYWTEPEFPERPERSKNQK 424
Qy	473 DSLNS-IEAPYGHQASIFSGRAFCEWRFAVAEMKAFLFVTLRRVQFBP 520
Db	425 DSIDPIYYTPPG-----TCGRNICIGMRFAFMNMKLIATIKVLQNSFKP 467

RESULT 3	
Q9C6S0	PRELIMINARY; PRT; 593 AA.
ID	Q9C6S0
AC	Q9C6S0;
DT	01-JUN-2001 (TREMELZel. 17, Created)
DT	01-JUN-2001 (TREMELZel. 17, Last sequence update)
DT	01-MAR-2004 (TREMELZel. 26, Last annotation update)
DE	Cytochrome P450, putative.
GN	Name=FSM6.19;
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsiis.
OX	NCBI_TaxID=3702;
RX	[1]
RP	SEQUENCE FROM N.A.
RA	Liu X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,



```

Db 356 -LASGDDVSSKQLRDDLTMTWLIAGHETSAAVLTWTFTYLLTTEPVSVAKLQEEVDVSGIDR 414
QY 376 MPTLDLDELNALPYLEAFVKSLSRLDPSPYANRECKLDEDFIPLAEPVIGRDSGVINEVRI 435
Db 415 FPTIQDKKLLKYTRVNSLSRLYPQPPVLIIRSI-----DNDILGEYPI 459
QY 436 TKGTWMLPLFNINRSGFYGEDAEFRPRWLEVD-----SLNSIEAPYGHQASFSIG 491
Db 460 KRGEIDFISVWNLHRSPLHWDAAEFNPERWELDPGPNENQNFSLPFG-----GG 512
QY 492 PRACFGWRFAVAEMKAPFLVTLRVOPE 519
Db 513 FRKICGDMFASPENVAIAMLIRFNFQ 540

RESULT 5
O97689 PRELIMINARY; PRT; 503 AA.
ID O97689
AC O97689;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cytochrome P450.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Schuetz E., Wieland E., Oellerich M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AF109068; AA004628.1; -.
DR HSSP; P14779; 1JPZ.
DR GO; GO:0005624; C:membrane fraction; ISS.
DR GO; GO:0004497; P:monooxygenase activity; ISS.
DR GO; GO:0006805; P:xenobiotic metabolism; ISS.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP4501.
DR InterPro; IPR008072; EP450_Cyp3A.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR01689; EP4501CYP3A.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 503 AA; 57267 MW; 22DB19F43C5988BB CRC64;

Query Match 14.6%; Score 417; DB 2; Length 503;
Best Local Similarity 27.9%; Pred. No. 1.9e-19;
Matches 160; Conservative 82; Mismatches 211; Indels 120; Gaps 24;

QY 12 GLAFAFNSAIA--FPSLYLAPRS-----SLYNLOGPNHTNYFTGNFLDILSARTGEH-- 63
Db 6 GFSTETWLLATSLVLLYLYGTYSHGLFKLGIPGPRPLPYF-GN-----ILGYRGVDHFD 61
QY 64 AKYREKYGTLRFAGIAGAPVNLSTDPKVFNVHM-KEAYDY-----PKPGMAARVLR 114
Db 62 KKCFOQYQKMGWFD-GRQPVLAITDPMITVLVKESYVFTNRSSFGPRGAMRT----- 116
QY 115 IATGQVVTAGEAHKRRHRRIMISLSAQAVKSWVPFLFKGMELVDKXMDAEKDMAV 174
Db 117 -----ALSLADEEMKRIITLLSTFTSGKLEKFFIISHYGDLLVSNLRKE-AEKG--- 167
QY 175 GESAGEKATRIETEGVDVKDVGCRATLDVNALAGFDYKDSL-----QNKTNELVY 226
Db 168 -----KPVTKQKIFGAYSNVDVITSTAFGWNTDFLANNPDPPFVNSKLLKF 213
QY 227 APVGLTDGFAPTLDSFKAIMWDVPVFTFMKRRHEIPIQTGLAVS-----RVGIELMEQKK 283
Db 214 SF-----FSLPIL--FPFPL-----TFILELVNLTLPFKSVNPFNRSIK 254
QY 284 QAVLGASDAQVDKQDQGRDILSLVRANITANLPESQKLSDBEVLAQISNLLFAGYET 343

```

```

Db 255 R-----MKESRLKDKQTHRVDFLQLMINSONSKETHTHKLSDLELVAQGVFFIAGYET 309
QY 344 SSTVLTWFMHRLSEDKAVQDKLREEICQI--DTDMFTLDELNALPYLEAFVKESRLDPP 401
Db 310 TSSLSLLVYELATHDPVQKQLEEDATFPSKALPSYDALAQMEYLDVMVNBILRLYPI 369
QY 402 SPYANRECLDKDEDFIPLAEPVIGRDSGVINEVRIKTGTWMLPLENINRSKFIYGSDAE 461
Db 370 AARLERYCKKDE-----IHGVSVPKGTVMVVFVSIHRDPELWPE--PEE 413
QY 462 FRPRWLEVDVTSLSNSTEAPYGHQASFSIGPRACFGWRFAVAEMKAPFLVTLRVOPEPI 521
Db 414 FRPERFSKKNKDSIN-----PYTV-LPFGTGPNCIGWRPALMNMKLLALVRVLQNFQSFKPC 468
QY 522 -----ISHPEYEHITLIISRPRIVG 541
Db 469 KETQTLKLSQGLIQEKPILLKWPDRDGTVG 501

RESULT 6
O97689 PRELIMINARY; PRT; 517 AA.
ID O97689
AC O97689;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative cytochrome P450.
GN Name=OJ1332_C12.1113;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AP003752; BAC10039.1; -.
DR HSSP; P14779; 1JPZ.
DR Gramene; O97689; 1.
DR GO; GO:0004497; P:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP4501.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 517 AA; 58038 MW; 3C3E35745026E138 CRC64;

Query Match 14.6%; Score 417; DB 2; Length 517;
Best Local Similarity 26.5%; Pred. No. 2e-19;
Matches 154; Conservative 95; Mismatches 202; Indels 130; Gaps 24;

QY 6 LITGALGAASFNSIASIAPFSLYLAP-----RRSSLYNLOGPNHTNYFTGNFLDILSARTGE 61
Db 7 MVAAAVAAVAVASWAPDAVVKLVWRPRAITRLRAQGVGGPGY-RFFSNLGEI--RRLRD 63
QY 62 EHA-----KYREKYGTLRFAGIAGAPVNLSTDPKVFNVHMKEAY 101
Db 64 EGAGVVDVSSDFFVPIVQPHFRKWIPLYGKTFMY-WFGARPTICLADVSMVRQLSDRT 122
QY 102 D-YPK---PGMAARVLRATGQVVTAGEAHKRRHRRIMISLSAQAVKSWVPFLFKGM 157
Db 123 GMYPKVNSNPFYPARLL-----GKGLVLTGDEWKRHKRVHFAFNMDKLKMMT----- 170
QY 158 ELVDKXMDAEKDMAVGESAGEKATRIETEGVDVKDVGCR-----ATLDVNALAGF--D 211
Db 171 -----VTMSDCAOSMISEWES-----ELGTGKDIVEIELSRPEELTADVISHTAFGSS 219

```

QY 212 YKSD-----SLQNKINELYAVFGLTDFAPTLDSFKALMWDFVFPYRTMKRRHEIP 263  
 Db 220 YKSGQVFLAOREIQ-----FLAF-----STFLSI-----QIP 247  
 QY 264 LTOGLAVSRVRYGIELMBOQKQAVLGASDAQVKKDV--OQRDILSLVLRANIAANLPES 321  
 Db 248 GSSYLPTKKNLKTWSVDKVRSMLTDIKSLRNKQVAGYGNLDLGLMLEA-CAPEHGES 306  
 QY 322 Q-KLSDBEVLAQISNLAFAGVETSTVLTWFMFRLSDKAVQDKLREIC-QIDTDMPTL 379  
 Db 307 QPQUSMEIITAECKTFFTFAGHDTTSHLLTWTMFLSLTHPEWQEKLEEVATECDGKVP 366  
 QY 380 DELNALPYLEAFVKESLRDPPSPYANRECLKDBDFPLAEPVIGRDGVSINVRITKGT 439  
 Db 367 DMLNKLKLVNMFLETLRLYGPVAFIQRRVNAELE-----LGGITVPEGT 411  
 QY 440 NVMPLPFLNRSKIYEDAEFPERPWLIEDVDTSLNSIEAPYGHQASFTSGPRACFGWR 499  
 Db 412 VLSIPATITHRDKVWGEDADIFKPERFKNGSVKAGYPNA-----LLSFSGPRACITQ 467  
 QY 500 FAVAEKAFVLTLLRRVQFEPILSHPEYEH--TLIISRRP 538  
 Db 468 FAMEAKAVAMILORFSF---TLSPKVVHVPTDVTILRPK 505

## RESULT 7

CP3T\_PIG ID CP3T\_PIG STANDARD; PRT; 503 AA.  
 AC P79401;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 03-JUL-2004 (Rel. 44, Last annotation update)  
 DE Cytochrome P450 3A29 (EC 1.14.14.1) (CYP11A29).  
 GN Name=CYP3A29;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=98347363; PubMed=9682441;  
 RT Nissen P.H., Winteroe A.K., Fredholm M.;  
 RT "Mapping of porcine genes belonging to two different cytochrome P450  
 subfamilies.";  
 RL Anim. Genet. 29:7-11(1998).  
 CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate  
 monooxygenases. In liver microsomes, this enzyme is involved in an  
 NADPH-dependent electron transport pathway. It oxidizes a variety  
 of structurally unrelated compounds, including steroids, fatty  
 acids, and xenobiotics.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.  
 CC -!- INDUCTION: P450 can be induced to high levels in liver and other  
 tissues by various foreign compounds, including drugs, pesticides,  
 and carcinogens.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z93099; CAB07513.1; -;  
 DR HSSP; P14779; 1JPZ.  
 DR GO; GO:0005624; C:membrane fraction; ISS.  
 DR GO; GO:0004497; F:monooxygenase activity; ISS.  
 DR GO; GO:0006805; P:xenobiotic metabolism; ISS.  
 DR InterPro; IPR001128; Cytochrome\_P450.

DR InterPro; IPR002401; EP450I.  
 DR InterPro; IPR008072; EP450\_CYP3A.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00463; EP450I.  
 DR PRINTS; PR01689; EP450IICYP3A.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;  
 KW Monooxygenase; Oxidoreductase.  
 FT METAL 442 442 Iron (heme axial ligand) (By similarity).  
 SQ SEQUENCE 503 AA; 57198 MW; E3D411B2674FD17F CRC64;

Query Match 14.5%; Score 414; DB 1; Length 503;  
 Best Local Similarity 27.4%; Pred. No. 3.1e-19;  
 Matches 158; Conservative 85; Mismatches 197; Indels 136; Gaps 23;

QY 12 GLAFAWSAIA--FFSLYLAPRS-----SLYNLOQPNHTNYFTGNFLDILSARTGEEH-- 63  
 Db 6 GFSTETWLLATSLVLLLYGTYSHGLFKLGIGCPRLPYF-GN--ILGYRKVDVDFD 61  
 QY 64 AKYREKYGSTLRPAGI--AGAPVLNSTDPKVFNVHM-KEAY-----DYPKPGMAARVL 113  
 Db 62 KKCFOQYG---KMWGVYDGRQPLLAVTDPNNIKSVLVKECYSVFTNRRSFGPLGARNAL 118  
 QY 114 RIATGDGVVTAEAGEAHRHRIMIPSLSAQAVKSMVPILFLEKGMELVDKMMEDAEKDVA 173  
 Db 119 SI-----AEDEWKRIRLLSPTFTSGKLEKMFPIISHYDGLLVSNLRK-AERG-- 167  
 QY 174 VGESAGEKKAIRLETEGVVDKDWGRATLDVMAAGFDYKSDSLQNK-----TNELYVAF 228  
 Db 168 -----KPTMKDIFGAVSMDVITSTAFGVNIDSLNNPQDPVENSKKLLK 212  
 QY 229 VCLTDGFAPTLDSPKAINWDF-----VPIYF-RTMKRRHEIPTLTOGLAVSRV 275  
 Db 213 PSFDFPFLSLTIFPFLPIPEVLNITLFPKSVNFTKSVKMKERLT----- 262  
 QY 276 TELMEQKQAVLGASDAQVDKDVQGRDILSLVRANIAANLPESQKLSDEEVLQISN 335  
 Db 263 ---DQCKRRV-----DLQLMINQSKEMDPHKSLSNEELVAQGI 301  
 QY 336 LLFAGYETSTVLTWFMFRLSDKAVQDKLREECQI--DTDMPTLDELNALPYLEAFVK 393  
 Db 302 FIFAGYETTSALLAYELATHDPVQOKLQEEIEATFPNKAPPTYDALLAQMEYLDVNVN 361  
 QY 394 ESLRLDPPSPYANRECLKDEDFIPLABPVIGRDGSVINEVRITKGTWVMLPFLNINSKF 453  
 Db 362 ETLLYPIAARLERACKQVE-----IHGVFVPKGVVVVFPVFLHRRDPD 406  
 QY 454 IYGDABEFPFRWLEDVDTSLNSIEAPYGHQASFTSGPRACFGWRFAVAKFAFLVTL 513  
 Db 407 LWPE-PBEFPRPERSKKHKTIN---PYTY-LPFGTGPRNCIGMRPALMMKLLALVRVL 460  
 QY 514 RRVOFEPI-----ISHPEYEHITLIISR 536  
 Db 461 QNFSFKCKETQIPLKLTUOGLTQPEKPVVILKILPR 496

## RESULT 8

CP39\_RAT ID CP39\_RAT STANDARD; PRT; 503 AA.  
 AC P51538; O64557; Q64631;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Cytochrome P450 3A9 (EC 1.14.14.1) (CYP11A9)  
 DE (3AH15).  
 GN Name=CYP3a9;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.



DR PROSITE; PS00086; CYTOCHROME P450; 1.  
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;  
KW Monooxygenase; Oxidoreductase.  
FT METAL 442 442 Iron (heme axial ligand) (By similarity).  
SQ SEQUENCE 503 AA; 57360 MW; 804223EAD0304238 CRC64;

	Query Match	14.3%; Score 408.5; DB 1; Length 503;
	Best Local Similarity	28.5%; Pred. No. 7.2e-19;
	Matches	165; Conservative 82; Mismatches 157; Indels 135; Gaps 27;
QY	7 LTGALGLAAFSWASIAFSLIAPRS-----SLYNLOPNHNTYFTGNFLDILSARTG-- 60	
DB	3 LIPSFSLETWTVLLAISLVLLYLTYSHGLFKLGVSGRPLPY-GN---VLRYRKVC 58	
QY	61 ----EEHAKYREKYSTLRPAGIAGAYLNSTDKPVNHVM-KRAYD-----YPXGM 108	
DB	59 EFDEECKTKYGOMGV----PEG--KQLLVITDPDVKTIVLKBCYSVFTRNRVRFGPMGI 113	
QY	109 AARVLRIATGDVVVTABGEAHRHRRIRIMISLSAQAVKMVPPIELEKGMELVDQMEDAA 168	
DB	114 MK-----NAVSAVEDQWKRTLTLPSTFTSGKLDMFFIIIGKYGDVLVRLRKE-A 164	
QY	169 EKDMAVGESAGEKKATLETEGYDDVKDWCGEATLDUMALAGDFYKSOSLONKITNELYAP 228	
DB	165 EK---GKS-----VNMKDI FGAYSMDVIITSTSGVINIDSLGNP----- 199	
QY	229 VGTDTGFAPT-----LDSP--KAIMWDF-VPYFR-----TMKRREHP-LTQGHAVS 271	
DB	200 ---QDPVENAKLLRFNIIDPFLLSVLFVLPFIPEVLNITMFPKSAVDFTK--SVK 254	
QY	272 RRVGIELMEOKQAVLGSASDAQVDDKVQGRDILSLVRANTAANLPESQKLSDEEVLA 331	
DB	255 RIRESRLKNQKPRV-----DFQLMINQSISKETDNHKALSDELMA 297	
QY	332 QINLLAFAGYTGSTVLTWMFHLSEDKAVODKURBEICOIDTDM-----PTDELNALP 386	
DB	298 QSUIFIFAGYETTNTLSFLXYLIATHPDVOOQLQES---IDATFPNKAPPTYDYVLAQME 354	
QY	387 YLFAFYKESLRLDPPSPYANRECLKBDFIPLAEIPVIGRGSVTNIEIKGWMLPLF 446	
DB	355 YLDMWNVNETLMRPPIAVRLDRCLKOVE-----IHGVSIFKGTAIVTVPIF 399	
QY	447 NINRSKFIVEDAEERPERPLEWDVTDLSNS-IEAPYGHQASTISGPACFGRFVADEM 505	
DB	400 VLHRDPQLWFE-PFERPERFSKXNDKSNIFYVLPPF-----TGPRNCIGMRFAIMNM 452	
QY	506 KAFLVTLRRVQEPFIISHPEYEHTLIISRPRIVGREK 544	
DB	453 KLAIVRLQNFSFKPC---KETQIELKINSOGILRPEK 487	

```

RESULT 10
QST485
ID Q6T485 PRELIMINARY; PRT; 514 AA.
AC Q6T485;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, last annotation update)
DE Cytochrome P450.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OC NCB:"TaxID=4565;
CX [1]
RN SEQUENCE FROM N.A.
RP Kong L., Anderson J.M., Ohm H.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RR -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC EMBL, AY437442; AAR11387.1; -.
CC InterPro; IPR001128; Cytochrome_p450.
CC InterPro; IPR002401; EP450I.
CC Pfam; PF00067; P450.1.
DR
DR
DR

```

```
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE      514 AA; 57578 MW; 576A11F12D25FDD8 CRC64;

Query March          14.2%; Score 405; DB 2; Length 514;
Best local similarity 25.3%; Pred. No. 1.3e-18;
Matches 146; Conservative 96; Mismatches 221; Indels 114; Gaps 21;

Qy   1 MFILVLITGALGLAAFSWASIAFPSSLYLAP-----RRSSLNLOQPNTNYFTGNFLDTLS 56
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy   1 MGLVWLVAAVVULASNAFNALVYLWRPRAITRQLRAQGVGGPGY-RFAGNLAIKQ 59
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy   57 ARTGEHHA-----KYREKYGSTLRFAGIAGAPVLNSTDPKVFNVMK 98
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy   60 LRADSAGAALDIGNHDFVPVRQPHFRKWIPIHGRTFLY-WFGARPSLCVADNVTVKQVLS 118
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy   99 EAYD-YEPKGWAARVLRIATGDGVVTAEGBAHKRHRIMIPSLPSAQAVKSNVPFILEKGM 157
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  119 DRSLYPKSGINPHIAL-LGKGLVLTGDGDWRKRHKRVHPAFNMWDKLKMT----- 169
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  158 ELVDKMEDAAEKDMAVGESAGSKKATRLBTEG---VDVKDWVGRAFTLVNALAGPDFYS 214
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  170 -----VTMSDCA-----GSNMSEWKKA-KMDKGSVEIDLSHQPEELTADVISHTAF---- 214
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  215 DSLQNKTNELYVA-----FVGLTDGPAPTLDSPKAIMWDVPVFRTRMKRHEIPLTGCLA 269
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  215 GSSYEQKKVFLAQRLEQFLAFSTVF-----NQIAPAFYLP 251
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  270 VSRVVGIELMEQKQAVLGSAISOAVDKDVQ--GRDIISLLVRANIATANLPESQK--L 324
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  252 TERNKIWKLDKEVRTMLMNIIGRLATKDIMGYGNDLLGLM----LEACAPEDRQNPLL 307
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  325 SDEVLQAQISNLIFAGVETTSVLTWNFHRLSDKAVQDKLREEIC-QIDTDMPTLDELN 383
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  308 SMDEIIDCKTPFFAGHDTSLLLTTWMFLLSTHPKWEKRLREVIRECGNGVPTGMDLN 367
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  384 ALPYLEAFVKESURLPPSPFYANRECLKDEDFIPLAEPVIGRDGVSINEVIRTKGTMMVL 443
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  368 KLQVNWNFLETURLYAPVASIQKAGSDLE-----VGGIKVPEGTVLTI 412
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  444 PLFNINRSFYIGEDAEEPRPERLEWDDVTDLSNLSIEAPYGHQSASFISGRPACFGWFAVA 503
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  413 PIATIHREDKEVGEDANEKPKVRPFENGVTGAKHPNA-----LLSFSSGSRSCIGNQPMFI 468
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  504 EMKAFVLTVLRVQFPFIISHFEYER--ITLIISRPR 538
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy  469 EAKAVTAVILQRFSPFS---LSPKYHAPMDVITLRPK 502
Db   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```

RESULT 11
ID Q6DV71 PRELIMINARY; PRT; 514 AA.
AC Q6DV71;
DT 01-OCT-2004 (TReMBLrel. 28, Created)
DT 01-OCT-2004 (TReMBLrel. 28, last sequence update)
DT 01-OCT-2004 (TReMBLrel. 28, last annotation update)
DE Cytochrome P450 CYP709C1.
GN Name=CYP709C1;
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RQ Kandel S., Werck-Reichhart D., Pinot F.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
SR EMBL; AY641449; AAT68297.1; -.
SQ SEQUENCE 514 AA; 57487 MW; 7B95CF7E520F701 CRC64;
Query Match 14.2%; Score 405; DB 2; Length 514;

```







-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54915; CAA3687.1; --  
 DR PIR; S14275; S14275.  
 DR HSSP; P14779; 1JPZ.  
 DR GO; GO:0005624; C:membrane fraction; ISS.  
 DR GO; GO:0004497; F:monooxygenase activity; ISS.  
 DR GO; GO:0006805; P: xenobiotic metabolism; ISS.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR InterPro; IPR002401; EP450I.  
 DR InterPro; IPR008072; EP450\_CYP3A.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00463; EP450I.  
 DR PRINTS; PR01689; EP450IICYP3A.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;  
 KW Monooxygenase; Oxidoreductase.  
 FT METAL 442 442 Iron (heme axial ligand) (By similarity).  
 SQ SEQUENCE 503 AA; 57721 MW; 52171D03F9BD5D87 CRC64;  
 Query Match 14.2%; Score 404; DB 1; Length 503;  
 Best Local Similarity 29.0%; Pred. No. 1.4e-18;  
 Matches 146; Conservative 83; Mismatches 184; Indels 90; Gaps 20;  
 QY 69 KYGSLRPFAGIAGPVLNSTDPKVFNVHM-KEAYDPKPGMAARVRIATG-----DGV 121  
 Db 67 KYGRMWGYPD-GRQVLAITDPMIKTVLKYCY-----SVFTNRRLTGPVGFKMSAI 118  
 QY 122 VTAEGEAKHRRIRIMIPSLSAQAVKSMVPIFLEKGMELVDKQMEDAAEKDVAEGESAGEK 181  
 Db 119 SLSEDEEWKMRLLSPTFTTGKLEMFPIIGYGVLVNNLRKE-AEKGA-----169  
 QY 182 KATLETGVVDVGRATLDVNALAGFDYKSDSLQNKTNELYAVFVGLTDCGAFPTLDS 241  
 Db 170 -----INLKDVFAGYSMDVITSTSGFNWIDSLNHPQD-----PFVEN 206  
 QY 242 FKAIM-WDFV-PVFTMKRRHEIPLTQGLA-VSRVGIEMLEOKKQAVLGASDQAVDK- 297  
 Db 207 TKLLKEDFDLPF-----FSILLPFLTPPEILNIMLPFKKVTDFRKSVRMKEGR 260  
 QY 298 -KDVQGR--DILSLVRANIAANLPESQKLSDEVLAQISNLLFAGYETSSVLTWMPHR 354  
 Db 261 LKDKQHRVDFLQLMINSQNSKEMDTHKALSDLELVAQSIIFIFAGYETTSLSFLMYE 320  
 QY 355 LSEDKAVODKLRREICQI--DTPMTLDLDELNALPYLEAFVKESLRDPPSPYANRECLD 412  
 Db 321 LATHPDVQKQBEIDATFFNKALPTVDALVQMEYLDVNLNETRLRYPIAGRLERYCKD 380  
 QY 413 EDPIAEPIVGRDGSVINEVRITKGMVNLPLFNINRSKIYGEDAEERFRMELEDVT 472  
 Db 381 VE-----ISGVFIPKGTVMVPTFLHRDQSLWPE-PEFRFRERSKKN 424  
 QY 473 DLSINSEAPYGHQASFIISGPRACFGWRFAVAEMKAFVTLRRVQPEPIISHPEYHITL 532  
 Db 425 DSIN-----PTY-LPFGTGRNCIGMRFAMNKLALVRVLQNFSEKPC-----KETQIDPL 475  
 QY 533 IISRPRIVREKGYQVRLQVCP 555  
 Db 476 KLNAGQIIQPEK---PIVLKVEP 495  
 RESULT 14  
 Q95L35  
 ID Q95L35  
 AC Q95L35;  
 PRELIMINARY; PRT; 503 AA.

DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Cytochrome P450 3A.  
 GN Name=CYP3A;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Soucek P., Zuber R., Anzenbacherova E., Linka M., Anzenbacher P.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Zuber R., Anzenbacherova E., Anzenbacher P.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 DR EMBL; AF424780; AAL1316.1; --  
 DR HSSP; P14779; 1JPZ.  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0016712; P:oxidoreductase activity, acting on paired d. . .; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR InterPro; IPR002401; EP450I.  
 DR InterPro; IPR008072; EP450\_CYP3A.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00463; EP450I.  
 DR PRINTS; PR01689; EP450IICYP3A.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 KW Heme; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 503 AA; 57085 MW; 5A17128CA9E50914 CRC64;  
 Query Match 14.1%; Score 403.5; DB 2; Length 503;  
 Best Local Similarity 27.8%; Pred. No. 1.6e-18;  
 Matches 152; Conservative 81; Mismatches 191; Indels 123; Gaps 22;  
 QY 12 GLAFAWSAIA--FFSLYLAPRRS-----SLYNLOGPNHTNYFTGNFILDLSARTGEH-- 63  
 Db 6 GFSTETVLLATSLVLLYGYTSHGLFKLGIQPRPLPYF-GN---ILGYRKGVDHFD 61  
 QY 64 AKYREKYGSLRFAGI--AGAPVLNSTDPKVFNVHM-KEAY-----DYPKPGVAARVL 113  
 Db 62 KKCQOYQ--KMGVDYDGRPLAVTDPMIKSVLVKECYSVFTNRRSFGPLGAMNAL 118  
 QY 114 RIATSDGVVTAEGEAHKEHRRIMIPSLSAQAVKSMVPIFLEKGMELVDKQMEDAAEKDMA 173  
 Db 119 SL-----AEDEWKRIRTLSPFTTSGKLEMFPIISHYGDLLVSNLRKE-AEKG-- 167  
 QY 174 VGESAGEKKAIRLETGVVDVGRATLDVNALAGFDYKSDSLQNK-----THLYVAF 228  
 Db 168 -----KPVTKMDIFGAYSMVDVITSTAFGVNIDSLNNPQDPPEVENSKKLLK 212  
 QY 229 VGLTDGFAPTLDSFKAINWDF-----VPYF-RTMKRRHEIPLTQGLAVSRVVG 275  
 Db 213 PSFPDPLLSLIFPFLTPPIPEVLNITLFPKSSVNVFTKSVKRMKESRLT----- 262  
 QY 276 IELMEQKQAVLGASDAQVDKQVGRDILSLVRANIAANLPESQKLSDEVLQAQISN 335  
 Db 263 ----DQKKRRY-----DLLQLMINSQNSKEMDPHKSLSNSEELVGGII 301  
 QY 336 LLFAGYETSSVLTWMTMFRHSEDKAVQDKLRREICQI--DTPMTLDLDELNALPYLEAFVK 393  
 Db 302 PIFAGYETSSALSLLAYELATHPDVQKQBEIEATFPNKAPTYPYDALAQMEYLDVNVN 361  
 QY 394 ESLRLDPPSPYANRECLKDEDFIFLABFVIGRDSVINEVRITKGMVNLPLFNINRSKFI 453  
 Db 362 ETLRLYPIAARLERACKKQVE-----IHGVFVPGKGTVMVPTFLHRDQSLWPE 406  
 QY 454 IYGEDAEERFRMELEDVTDSLNSIEAPYGHQASFIISGPRACFGWRFAVAEMKAFVTL 513  
 Db 407 LWPE-PEFRFRERSKKNKOTIN-----PTY-LPFGTGRNCIGMRFAMNKLALVKVL 460

```

QY 514 RRQVFP 520
DB 461 QNFSKP 467

RESULT 15
Q9LUD2
ID Q9LUD2 PRELIMINARY; PRT; 515 AA.
AC Q9LUD2;
DT 01-OCT-2000 (TREMREL.15, Created)
DT 01-OCT-2000 (TREMREL.15, Last sequence update)
DT 05-JUL-2004 (TREMREL.27, Last annotation update)
DE Cytochrome P450 (AT3G14620/MIE1.12).
OS Arabidopsis thaliana (Mouse-ear Cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RP MEDLINE=20277480; PubMed=10819329;
RX Nakamura Y.;
RA "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
RN (2)
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi W., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AB023038; BAB02394.1; -
DR EMBL; AY052208; AAK97679.1; -
DR HSSP; P14779; 10PZ.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP4501.
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 515 AA; 58643 MW; 9F4EFCF7686F55A1 CRC64;

Query Match 14.1%; Score 403; DB 2; Length 515;
Best Local Similarity 26.5%; Pred. No. 1.7e-18;
Matches 134; Conservative 87; Mismatches 191; Indels 94; Gaps 19;

QY 42 NHTNYFTGNFLDILSARTGEHAKYREKYSTLFAAGAPVLNSTDPKVFHVMKEAY 101
DB 76 NLTDYTHRVMPLIQ-QTVKHGK-----TSYMMWGPIASVIV---TKPEHIKDLNRY 126

QY 102 DYPKPGMAARVLRATGCVTAGEAHKRHRIMIPSLSAQAVKSMVPIFLEKGMELVD 161
DB 127 DFPKPPVHPHIVELFAT--GVALYGEKWKSRKRIINFSFLEKUKIMIPAFYSCSEMIS 184

QY 162 KXMEDAAEKDVAVGESAGEKKATRLTEGVVDKDWGVRATLDVNALAGF--DYKSD---- 215
DB 185 KW-----EKLVTQSSNE-----IDWPFVLGDLTSDVISRTAFGSSVEEGKRI 229

```

```

QY 216 SLQKNTNELYVAFVGLTDFGAPTLDSEKAIWMDVFPYFRIMKRRHEIPLTQGLAVSRVVG 275
DB 230 ELQEEQGRRLV-----KALELAFIPQMRFLPTKNNLRMQ---INKEVK 270
QY 276 IELME--QKKQAVLGSASDAQVKDVQGRDILSLVRANIAANLPESQKLSDEEVLQAI 333
DB 271 SRLREIIMKQ-----RQMTGEAPKNDLGLLESNSG-----DHGMSIEDVVEEC 317
QY 334 SNLLFAGYETSTVLTWMFRLSDEKAVQKREEIFCOI--DTDMFTLDELNALPYLEAF 391
DB 318 RLPHFAGQETTAVLLVMTIMLSHHQWQDQAEELIKVICKNNKPNFDALSCLKTMSMI 377
QY 392 VKESIRLDPPSPYANRECLKDEDFIPLAEPVIGDGSVINEVRITKGTVMMLPLFNINRS 451
DB 378 LNEVRLYPPGILLGRIVEK-----ETKLGED-----MTLPGGAQVVIPLVMVHRD 423
QY 452 KFIYGEDAEERPERMLEDVTDLSLNSIEAPYHQASFIS---GPRACFGWRFAVENKAF 508
DB 424 PELWGEDVHEFNPERFADGISKATK-----NQVSFLPFGWGPFCPCGQNFALMEAKMA 476
QY 509 LFTVLRVQFE--PIISHPEYEHITL 532
DB 477 LVLLQRFSPFELSPSYTHAPHTVLT 502

```

Search completed: December 14, 2004, 21:47:10  
Job time : 225 secs